CHINA'S MACROECONOMIC TIME SERIES: METHODS AND IMPLICATIONS
VERY PRELIMINARY DRAFT

PATRICK HIGGINS AND TAO ZHA

Abstract. We show how to construct a core set of macroeconomic time series usable for studying China's macroeconomy systematically. When applicable, we document our construction methods in comparison to other existing approaches.

Date: July 20, 2015.

Key words and phrases. Interpolation methods, core set, impulse responses, minimized errors.

JEL classification: E01, E2, E3, E4.

This research is supported in part by the National Natural Science Foundation of China Project Numbers 71473168 and 71473169. The views expressed herein are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Atlanta, the Federal Reserve System, or the National Bureau of Economic Research.

I. Introduction

This paper contains a detailed description of methods for constructing a core of China's macroeconomic time series usable for empirical studies. Some of these series have been used in Chang, Chen, Waggoner, and Zha (2015). Much of the raw source data for this study comes from the CEIC (China Economic Information Center, which belongs to Euromoney Institutional Investor Company). A variable name in this paper is referred to in bold face.

The notation $X_{t,q}$ denotes the value of X in the qth quarter of year t. $X_{t,m}$ denotes the value of X in the mth month of year t. We adopt the notational convention that $... = X_{t-2,q+8} = X_{t-1,q+4} = X_{t,q} = X_{t+1,q-4} = X_{t+2,q-8}...$, and $... = X_{t-2,m+24} = X_{t-1,m+12} = X_{t,m} = X_{t+1,m-12} = X_{t+2,m-24}...$ Likewise, X_t denotes the annual total of X in year t. Common abbreviations we use are "PY=100" for an index measuring 100 times the gross growth rate of an annual series, "PYm=100" for an index measuring 100 times the gross growth rate since the same month of the prior year, "PYq=100" for an index measuring 100 times the gross growth rate since the same quarter of the prior year, and "PM=100" for 100 times the gross monthly growth rate. The abbreviation "ytd" stands for "year-to-date" while "yoy" means "year-over-year" or the growth rate from the same period of the previous year. 100 times the gross growth rate is called a "growth index". SA stands for seasonally seasonally adjusted.

II. Construction methods

II.1. **Measures of value-added and expenditure GDP.** The two measures of gross domestic product (GDP) that we treat are GDP measured by the value-added approach (GDP-va) and GDP measured by the expenditure approach (GDP-exp) as the sum of consumption, gross domestic capital formation, government spending and net exports. Quarterly data on GDP-exp and its subcomponents are not published by China's National Bureau of Statistics (NBS) and are constructed described later in this section. Our algorithm for constructing quarterly real GDP-va – **RGDP_va_Q** – is:

Step 1: Annual nominal GDP-va – **NGDP_va_A** – is the CEIC ticker CATA. Annual real GDP-va – **RGDP_va_A** – is set to **NGDP_va_A** in 2008 and cumulated forwards and backwards with the PY=100 index of real GDP-va **RGDP_PY_A** (CEIC ticker CAOB). **RGDP_yoy_Q** [CEIC ticker CATA] is real GDP-va "PYq=100" while **RGDP_ytd_Q** is "ytdPYq=100".

Step 2: Given a guess for $Y^{\text{real},2008RMB}_{2008,1}$ [**RGDP_va_Q**] we find the unique set of values of $Y^{\text{real},2008RMB}_{2008,2}$, $Y^{\text{real},2008RMB}_{2008,3}$, $Y^{\text{real},2008RMB}_{2008,4}$, $Y^{\text{real},2008RMB}_{2009,1}$, $Y^{\text{real},2008RMB}_{2009,2}$, $Y^{\text{real},2008RMB}_{2009,3}$,and $Y^{\text{real},2008RMB}_{2009,4}$

that are mutually consistent with the four quarterly 2009 values of **RGDP_yoy_Q** and **RGDP_ytd_Q**. After solving for these seven unknowns, we verify that

$$Y_{t,q}^{\text{real},2008RMB} = \sum_{q=1}^{4} Y_{t-1,q}^{\text{real},2008RMB} \tag{1}$$

Obviously (1) will not hold for the first guess $Y_{2008,1}^{\text{real},2008RMB}$; by using the MatLab function fzero.m, the unique value for $Y_{2008,1}^{\text{real},2008RMB}$ that satisfies (1) is easily found.

Step 3: We use the 2009 quarterly values of **RGDP_va_Q** and the 2010q1-present quarterly values of **RGDP_ytd_Q** to get the 2010q1-present values of $Y_{t,q}^{\text{real},LEV}$ by means of the equation[1]:

$$Y_{t,q}^{\text{real},2008RMB} = (\sum_{h=1}^{q} Y_{t-1,h}^{\text{real},2008RMB})\frac{Y_{t,q}^{\text{real},ytdPY=100}}{100} - (\sum_{h=1}^{q-1} Y_{t,h}^{\text{real},2008RMB}) \tag{2}$$

Step 4: We use the 1999q4-2008q4 values of **RGDP_yoy_Q** and the 2008 quarterly values of **RGDP_va_Q** to get the 1998q4-2007q4 values of $Y_{t,q}^{\text{real},LEV}$ by means of the equation:

$$Y_{t-1,q}^{\text{real},LEV} = 100\frac{Y_{t,q}^{\text{real},LEV}}{Y_{t,q}^{\text{real},yoyPY=100}} \tag{3}$$

Step 5: Use the 1992q1-1999q3 values of **RGDP_ytd_Q** and the 1999q1-1999q3 quarterly values of **RGDP_va_Q** to get the 1991q1-1998q3 values of $Y_{t,q}^{\text{real},LEV}$ by means of the equation:

$$Y_{t,q}^{\text{real},2008RMB} = \frac{100}{Y_{t+1,q}^{\text{real},ytdPY=100}}(\sum_{h=1}^{q} Y_{t+1,h}^{\text{real},2008RMB}) - (\sum_{h=1}^{q-1} Y_{t,h}^{\text{real},2008RMB}) \tag{4}$$

A warning is in order. Except for the base year (chosen to be 2008) and the year following the base year (2009), in principle, one could choose either $Y_{t,q}^{\text{real},ytdPY=100}$ or $Y_{t,q}^{\text{real},yoyPY=100}$ to solve for $Y_{t,q}^{\text{real},2008RMB}$ in steps 3-5. We made the choices we did based on the decimal precision of the data; i.e. whenever unrounded data is available we use that. However, this strategy would be sub-optimal if the published series $Y_{t,q}^{\text{real},yoyPY=100}$ and $Y_{t+1,q}^{\text{real},ytdPY=100}$ are not consistent with each other due to the treatment of revisions. Haver Analytics' data description for their seasonally adjusted real GDP-va series[2] suggests that differences due to

[1]We adopt the notational convention that $\sum_{h=1}^{0} X_{t,h} \equiv 0$ for any quarterly time series X.

[2]According to Haver's documentation for this series: "China's quarterly real GDP in billions of 2000 yuan is calculated by Haver Analytics by taking the nominal values of the 4 quarters of 2000 and growing them forward using the 12-month percent changes of year-to- date GDP in 1990 prices published by the National Bureau of Statistics. Data are seasonally adjusted by Haver Analytics using X12-ARIMA. China's National

the treatment of revisions could indeed be an issue. Holz (2013c,b) make related comments about revisions to annual GDP. For example, page 8 of Holz (2013c) states "*Every year, the NBS in the Statistical Yearbook provides a set of revised nominal GDP and sectoral VA for the last previously published annual data (and similarly for expenditures)...* ***Real growth rates are typically not revised****... Throughout the early 1990s, first and later published real GDP growth rates differ by up to approximately one percentage point. Between 1995 and 2003, real growth rates were rarely revised.*" One necessary condition for **RGDP_yoy_Q** and **RGDP_ytd_Q** to be mutually consistent is that $Y_{t,1}^{\text{real},ytdPY=100} = Y_{t,1}^{\text{real},yoyPY=100}$ for all t. This equality, does indeed hold. A second, less robust, check uses the approximation

$$\log\left(\frac{\sum_{q=1}^{4} X_{t,q}}{\sum_{q=1}^{4} X_{t-1,q}}\right) \approx \frac{1}{4}\sum_{q=1}^{4}\log\left(\frac{X_{t,q}}{X_{t-1,q}}\right) \tag{5}$$

for a variable $X_{t,q}$ in levels [like nominal or real GDP]. The less volatile $\log(\frac{X_{t,q}}{X_{t-1,q}})$, the better the approximation. If (5) does not hold approximately, it may suggest inconsistency between two measures of a level. Figure 1 plots the real GDP discrepancy

$$100[\{\frac{1}{4}\sum_{q=1}^{4}\log(\frac{Y_{t,q}^{\text{real},yoyPY=100}}{100})\} - \log(\frac{Y_{t}^{\text{real},ytdPY=100}}{100})]$$

and the nominal GDP discrepancy

$$100[\{\frac{1}{4}\sum_{q=1}^{4}\log(\frac{Y_{t,q}^{nom,RMB}}{Y_{t-1,q}^{nom,RMB}})\} - \log(\frac{Y_{t}^{nom,RMB}}{Y_{t}^{nom,RMB}})]$$

The nominal GDP discrepancy is plotted since it is entirely accounted for by the approximation error in (5), as $\sum_{q=1}^{4} Y_{t,q}^{nom,RMB} = Y_{t}^{nom,RMB}$. The real discrepancy can be much larger than the nominal discrepancy, suggesting some possible inconsistencies between CEIC's time series for $Y_{t,q}^{\text{real},yoyPY=100}$ and $Y_{t}^{\text{real},ytdPY=100}$.

Putting this issue aside, we seasonally adjust the logarithms of **NGDP_va_Q** and **RGDP_va_Q** thereby obtaining the series **NGDP_va_Q_SA** and **RGDP_va_Q_SA**. The SAS procedure

---

Bureau of Statistics publishes two sets of quarterly growth rates for China's real GDP: 1) 12-month percent changes of GDP in 1990 prices. 2) 12-month percent changes of year-to-date GDP in 1990 prices. *Haver Analytics uses 12-month percent changes of year-to-date GDP in 1990 prices in calculating China's quarterly real GDP in billions of 2000 yuan because these growth rates contain revisions while the 12-month percent changes of GDP in 1990 prices do not.*"

"proc X12" is used to do the seasonal adjustment. The automdl option is used to select the values of p, d, q, p_s, d_s, and q_s for the seasonal ARIMA(p,d,q)(p_s,d_s,q_s) model where the log of the seasonally adjusted series is an ARIMA(p,d,q) process and the log of the seasonal factors is a seasonal-ARIMA(p_s,d_s,q_s) process. Unlike the U.S. *National Income and Product Accounts* (NIPA), the sum of the quarterly seasonally adjusted series cannot be restricted to equal the annual total in the seasonal adjustment routine[3]. We do a final clean-up step to enforce $\sum_{q=1}^{4} Y_{t,q}^{nomSA,RMB} = Y_t^{nom,RMB}$ and $\sum_{q=1}^{4} Y_{t,q}^{\mathrm{real}\,SA,2008RMB} = Y_t^{\mathrm{real},2008RMB}$ for all years t. The adjustment is done using the adapted interpolation method of Fernandez (1981) described in Appendix A. As with the U.S. NIPA, the quarterly implicit GDP price deflator **DGDP_va_Q_SA** is the ratio of **NGDP_va_Q_SA** to **RGDP_va_Q_SA**. Figure 2 shows the growth rate of our series is compared with alternative measures constructed by three sources: Haver Analytics, the Federal Reserve Board, and China's National Bureau of Statistics (NBS). The Federal Reserve series was downloaded from their FAME database. The series from the NBS is seasonally adjusted by that source agency and starts in 2010q4. We see there are some differences in all of the series. The Federal Reserve/FAME series and Haver's series appear to be the closest each other, although there is still a notable difference in those growth rates in 2003q1 and 2003q2. Although it is difficult to say which series is "best", one check we can do is to convert the quarterly series to annual real GDP-va and compare the constructed real annual growth rates with the published real annual growth rate from the NBS [taken from the CEIC]. Figure 3 has this comparison. By construction, the growth rates derived from the aggregated **RGDP_va_Q_SA** and **RGDP_PY_A** variables are identical. The differences for the other series are fairly small except for the Federal Reserve/FAME series in the early- to mid- 2000s.

We interpolate annual nominal GDP-exp – **NGDP_A** – with **NGDP_va_Q_SA** using Fernandez (1981)'s adapted method described in Appendix A. The interpolated real expenditure series – **RGDP_exp_Q_SA** – is deflated by dividing the nominal interpolated series (**NGDP_exp_Q_SA**) by the implicit GDP deflator (**DGDP_va_Q_SA**). Figure 4 compares the growth rates of **RGDP_va_Q_SA** and **RGDP_exp_Q_SA**. Figure 5 plots the HP-filtered cyclical components using both the quarterly data and annual data for 1992-2013[4]. The quarterly cycles appear reasonable relative to the annual cycles.

[3]This is because we are seasonally adjusting the logarithm of the series. If we were seasonally adjusting the level, the restriction could be imposed by proc X12.

[4]Following Ravn and Uhlig (2002) we set the smoothing parameter $\lambda = 6.25$ for the annual series so that it is nearly comparable to $\lambda = 1600$ parameter for the quarterly series.

II.2. **Components of expenditure GDP.** The notation $X_{t,q,m}$ denotes the value of X in the mth month of the qth quarter of year t. For example, $X_{2004,2,3}$, denotes June 2004. To preview our strategy, if we write nominal quarterly GDP-exp (SA) as

$$\begin{aligned} Y_{t,q}^{EXP-nomSA,RMB} &= C_{t,q}^{nomSA,RMB} + IFixed_{t,q}^{nomSA,RMB} + \\ &\quad V_{t,q}^{nomSA,RMB} + G_{t,q}^{nomSA,RMB} + NX_{t,q}^{nomSA,RMB} \end{aligned} \tag{6}$$

our approach will be to interpolate all of the components of quarterly GDP-exp (SA) except for change in inventories $[V_{t,q}^{nomSA,RMB}]$. Interpolated $V_{t,q}^{nomSA,RMB}$ will be derived as a residual since it is the only component that does not have a natural interpolater. Our full algorithm is the following.

Step 1: *Clean monthly fixed asset investment.* The monthly ytd series of fixed assets investment (FAI) **NFAInv_ytd_M** (CEIC ticker COBDJU) has two subcomponents which we will utilize: "capital construction" (**NFAInv_CP_ytd_M** – CEIC ticker COCA) and "innovation" (**NFAInv_INN_ytd_M** – CEIC ticker CODA). These are easily converted to quarterly averages using the formulas.

$$X_{t,1} = X_{t,1,3}^{ytd} \tag{7}$$

$$X_{t,q} = X_{t,q,3}^{ytd} - X_{t,q-1,3}^{ytd}(2 \le q \le 4) \tag{8}$$

After undoing this year-to-date accumulation, we define **NFAInvCPINN_Q** as the sum of the quarterly values of "capital construction" and "innovation" FAI. Figure 6 plots $100\log(\frac{X_{t-1,q}}{X_{t,q}})$ – the negative of the 4-quarter logarithmic percent change lagged 4-quarters – for total quarterly FAI [**NFAInv_Q**] and "FAI-Capital Construction+Innovation" [**NFAInvCPINN_Q**]. We see there is a gigantic outlier in **NFAInv_Q** in 1994q4. Perhaps FAI was genuinely very low in 1994q4 and the outlier is not due to a break in the data. If this were the case, we might expect full year "gross fixed capital formation" [**NGFCF_A** – CEIC ticker CALCA] growth in 1995 to be very strong as the 1995q4 would look very strong relative to 1994q4. But this is not the case, as we see in figure 7. In 1995, growth in gross fixed capital formation more closely resembles growth in "FAI-Capital Construction+Innovation" than it does growth in total FAI. Therefore, we presume that the low 1994q4 value for **NFAInv_Q** is an outlier and adjust it with the formula

$$
\begin{aligned}
NFA_{1994,4}^{totAdj,RMB} \quad = \quad & NFA_{1995,4}^{tot,RMB}\exp(\frac{1}{2}[\{\log(\frac{NFA_{1994,3}^{tot,RMB}}{NFA_{1995,3}^{tot,RMB}})+ \\
& \log(\frac{NFA_{1995,1}^{tot,RMB}}{NFA_{1996,1}^{tot,RMB}}))\}+\{\log(\frac{NFA_{1994,4}^{CC+Inn,RMB}}{NFA_{1995,4}^{CC+Inn,RMB}})- \\
& [\log(\frac{NFA_{1994,3}^{CC+Inn,RMB}}{NFA_{1995,3}^{CC+Inn,RMB}})+\log(\frac{NFA_{1995,1}^{CC+Inn,RMB}}{NFA_{1996,1}^{CC+Inn,RMB}})]\}])
\end{aligned} \tag{9}
$$

For all other values besides 1994q4, the outlier adjusted series **NFAInv_Qadj** is set to the unadjusted series **NFAInv_Q** [i.e. $NFA_{t,q}^{totAdj,RMB} = NFA_{t,q}^{tot,RMB}$ for $t \neq 1994$ or $q \neq 4$]. The negative of the 4-quarter logarithmic percent change lagged 4-quarters of the adjusted series is also plotted in figure 6. We see that the outlier is eliminated.

Step 2: *Aggregate monthly retail consumption and government expenditures.* Monthly retail sales of consumer goods **NRetC_M** is CEIC ticker CHBA while the ytd measure **NRetC_ytd_M** is CEIC ticker CHBE. To define the quarterly level of retail sales **NRetC_Q**, we use equations (7) and (8) for $t \geq 1994$ with **NRetC_ytd_M**. For $t < 1994$, we use $NRC_{t,q}^{lev,RMB} = \sum_{m=1}^{3} NRC_{t,q,m}^{lev,RMB}$ to set **NRetC_Q** from **NRetC_M**. Quarterly government expenditures from the Mininstry of Finance **NGovExp_FM_Q** are aggregated from the monthly CEIC series CFPAB.

Step 3: We convert dollar values of monthly goods exports [**NEXP_USD_M** – CEIC ticker CJAA] and monthly goods imports [**NIMP_USD_M** – CEIC ticker CJAC] from dollars to RMB. The CEIC's monthly exchange rate series **FX_avg_M** (CEIC ticker CMDAA) starts in 1994m1, so we backcast it using the exchange rate series from the Federal Reserve's G.5 statistical release and the ratio splice backcasting method described in Appendix B. The spliced exchanged rate series is multiplied by dollar exports/imports to get renminbi exports/imports and aggregated to the quarterly frequency [**NEXP_RMB_Q** and **NIMP_RMB_Q**]. The trade balance is **NNetEXP_RMB_Q** = **NEXP_RMB_Q** - **NIMP_RMB_Q**. Figure 8 plots net exports and the annual growth rates of the components of GDP and their interpolaters (prior to seasonal adjustment).

Step 4: *Seasonally adjust data that have been cleaned and aggregated from monthly to quarterly.* The 6 series that we seasonally adjust are the logarithms of **NRetC_Q**, **NEXP_RMB_Q**, **NIMP_RMB_Q**, **NFAInvCPINN_Q**, **NFAInv_Qadj**, and **NGovExp_FM_Q**. The seasonal adjustment algorithm is implemented exactly as it was for nominal and real GDP-va. The seasonally adjusted versions of all of these variables all have "_SA" appended to the end of their name; e.g. seasonally adjusted **NRetC_Q** is **NRetC_Q_SA**.

Step 5: *Interpolate nominal gross fixed capital formation.* As we see in figure 8, the annual averages of the series "total FAI" [**NFAInv_Qadj**] and "FAI: Capital Construction+Innovation" [**NFAInvCPINN_Q**] have very similar growth rates over their common sample period. Therefore, we backcast "total FAI, SA" [**NFAInv_Qadj_SA**] with "FAI: Capital Construction+Innovation, SA" [**NFAInvCPINN_Q_SA**] using the ratio splice backcasting method described in Appendix B. The spliced series – **NFAInv_QadjSplice_SA** – is used to interpolate nominal gross fixed capital formation [**NGFCF_A** – CEIC ticker CALCA] using the adapted interpolation method of Fernandez (1981) described in Appendix A. The seasonally adjusted series is **NGFCF_Q_SA** ($IFixed_{t,q}^{nomSA,RMB}$).

Step 6: *Interpolate nominal consumption.* Seasonally adjusted retail goods consumption [**NRetC_Q_SA**] is used to interpolate total nominal consumption [**NC_A** – CEIC ticker CALBA] using the adapted interpolation method of Fernandez (1981) described in Appendix A. The seasonally adjusted series is **NC_Q_SA** ($C_{t,q}^{nomSA,RMB}$).

Step 7: *Interpolate government spending.* Since the interpolater government expenditure series from the Ministry of Finance [**NGovExp_FM_Q_SA**] starts in 1995q1, we need to backcast this series. We do this by estimating the regression

$$\begin{aligned}\Delta\log(G_{t,q}^{MOF,nomSA}) \quad = \quad & [1, \Delta\log(G_{t,q+1}^{MOF,nomSA}), \Delta\log(G_{t,q+2}^{MOF,nomSA}) \\ & , \Delta\log(Y_{t,q}^{EXP-nomSA,RMB}), \Delta\log(C_{t,q}^{nomSA,RMB}), \\ & \Delta\log(IFixed_{t,q}^{nomSA,RMB})]\boldsymbol{\beta}+u_t^g\end{aligned} \tag{10}$$

Because we are backcasting, the two AR terms are leads rather than lags. Contemporaneous growth rates of GDP-exp, consumption, and gross capital formation are included because of the a-priori view that they should be informative about government spending growth due to the national income identity (6). Backcasting with (10) generates values $\Delta\log(\hat{G}_{t,q}^{MOF,nomSA})$ back to 1992q2. To backcast back to 1990q1, we pad the series $\Delta\log(\hat{G}_{t,q}^{MOF,nomSA})$ with backcasts from the AR(2) model with leads

$$\Delta\log(G_{t,q}^{MOF,nomSA}) = [1, \Delta\log(G_{t,q+1}^{MOF,nomSA}), \Delta\log(G_{t,q+2}^{MOF,nomSA})]\boldsymbol{\beta}+u_t^g \tag{11}$$

The series $G_{t,q}^{MOF,nomSA}$ padded with backcasts through 1989q4 is used to interpolate the government spending GDP component [**NGovExp_A** – CEIC ticker CALBD] with the adapted interpolation method of Fernandez (1981) described in Appendix A. The seasonally adjusted series is **NGovExp_Q_SA** ($G_{t,q}^{nomSA,RMB}$).

Step 8: *Interpolate net exports.* The seasonally adjusted goods trade balance [**TradeBal_Q_SA**] is defined as the difference between seasonally adjusted goods exports [**NEXP_RMB_Q_SA**] and seasonally adjusted goods imports [**NIMP_RMB_Q_SA**]. The same convention is used in the U.S. NIPAs as well. Net exports are somewhat tricky to interpolate as they cannot be converted to logs so that Fernandez (1981)'s interpolation will not work. Since the variance of nominal net exports grows over time, Chow and Lin (1971) does not seem to be a good interpolation approach either. Our strategy is to take nominal annual GDP-exp [**NGDP_A**] and convert it to quarterly values by assigning one-fourth of the annual value to each quarter of the year. Call this series **repNGDP_A_Q**. The seasonally adjusted goods trade balance as a percentage of GDP is defined as **TradeBal_Q_SA** = 100***NNetEXP_RMB_Q_SA**/(1000***repNGDP_A_Q**)[5]. Total annual net exports as a percentage of GDP is defined as **TradeBal_A** = 100***NNetExp_A/NGDP_A**, where **NNetExp_A** (CEIC ticker CALD) is net exports in the national income identity (6). We create the series **TradeBal_Q_Denton** by interpolating **TradeBal_A** with **TradeBal_Q_SA** using first-difference Denton interpolation[6]. These series are plotted in figure 9. Seasonally adjusted net exports [$NX_{t,q}^{nomSA,RMB}$] are defined as **NNetExp_Q_SA_Denton** = (**TradeBal_Q_Denton*****repNGDP_A_Q**)/100.

Step 9: We have interpolated nominal GDP-exp and all of its components except for inventory investment. Interpolated inventory investment **NInvty_Q_SA** [$V_{t,q}^{nomSA,RMB}$] is defined as the residual.

$$\begin{aligned} V_{t,q}^{nomSA,RMB} &= Y_{t,q}^{EXP-nomSA,RMB} - (C_{t,q}^{nomSA,RMB} + \\ & \quad IFixed_{t,q}^{nomSA,RMB} + G_{t,q}^{nomSA,RMB} + NX_{t,q}^{nomSA,RMB}) \end{aligned} \tag{12}$$

The series is plotted as percentage of nominal GDP-exp (SA) in figure 10. It is somewhat noisy, but there do not appear to be any large outliers. Seasonally adjusted nominal gross capital formation **NGCF_Q_SA** [$IGross_{t,q}^{nomSA,RMB}$] is defined as

$$IGross_{t,q}^{nomSA,RMB} = IFixed_{t,q}^{nomSA,RMB} + V_{t,q}^{nomSA,RMB} \tag{13}$$

The nominal series are converted to real by dividing by the implicit GDP deflator [**DGDP_va_Q_SA**]. The HP filtered cycles of the subcomponents of real GDP are plotted in figure 11.

[5]1000 is needed in denomanaitor since net exports of goods are in millions of yuan and GDP is in billions of yuan.

[6]The function call in Matlab is essentially denton_uni(TradeBal_A,TradeBal_Q_SA,2,1,4). This is modified slightly to handle the missing values that the original function does not allow. The Matlab toolbox with Denton and other interpolation routines is available at http://www.mathworks.com/matlabcentral/fileexchange/24438-temporal-disaggregation-library

II.3. **Decomposing investment by sector.** We decompose quarterly nominal gross fixed capital formation (SA) into the following five sectors: state-owned enterprises (SOE) excluding government, "private" enterprises, other non-SOE enterprises, households, and government. CEIC has annual "gross fixed capital formation" (GFCF) for 1992-2012 decomposed as in table 1 where CAAFTQ = (CAAFWG + CAAFYW + CAAGBM + CAAGEC). We directly interpolate the household sector's share of total GFCF – $\frac{CAAGEC}{CAAFTQ}$ – and the government's share of total GFCF – $\frac{CAAGEC}{CAAFTQ}$ – from the annual frequency to the quarterly frequency with Chow-Lin interpolation[7]. The household share $\frac{CAAGEC}{CAAFTQ}$ is interpolated with seasonally adjusted retail sales as a share of seasonally adjusted nominal GDP-exp; $\frac{\textbf{NRetC\_Q\_SA}}{\textbf{NGDP\_exp\_Q\_SA}}$. As seen in figure 12, the movements in $\frac{CAAGEC}{CAAFTQ}$ broadly follow the movements in $\frac{\textbf{NRetC\_Q\_SA}}{\textbf{NGDP\_exp\_Q\_SA}}$. Chow-Lin interpolation estimates

$$\frac{CAAGEC}{CAAFTQ}_{t,q} = [1, \frac{\textbf{NRetC\_Q\_SA}}{\textbf{NGDP\_exp\_Q\_SA}}_{t,q}]\boldsymbol{\beta} + e_{t,q} \tag{14}$$

where $e_{t,q}$ is an AR(1) error process[8]. As $\frac{CAAGEC}{CAAFTQ}_{t,q}$ is missing beyond 2012q4, we can project it forward using (14), future values of $\frac{NRetC\_Q\_SA}{NGDP\_exp\_Q\_SA}_{t,q}$, and projections of $e_{t,q}$ based on $\hat{e}_{t=2012,q=4}$ and the AR(1) error coefficients. The projections are also shown in figure 12. We interpolate the government's share of total GFCF – $\frac{CAAGBC}{CAAFTQ}$ – using $\frac{\textbf{NGovExp\_FM\_Q\_SA}-\textbf{NGovExp\_Q\_SA}}{\textbf{NGDP\_exp\_Q\_SA}}_{t,q}$ which is "Quarterly Government Expenditure (from Ministry of Finance), SA" minus "Interpolated Nominal NIPA Government Expenditure, Quarterly SA" as a share of GDP-exp. The rationale is that government investment should be related to the difference of government revenue and government consumption. Figure 13 shows the interpolation; the correspondence between the interpolater $\frac{\textbf{NGovExp\_FM\_Q\_SA}-\textbf{NGovExp\_Q\_SA}}{\textbf{NGDP\_exp\_Q\_SA}}_{t,q}$ and the annual share $\frac{CAAGEC}{CAAFTQ}$ is much less tight than it is for the household share of GFCF[9]. Total enterprise investment's share of GFCF is backed out as 1 minus the sum of the household and government share.

$$\frac{CAAFWG + CAAFYW}{CAAFTQ}_{t,q} = 1 - \frac{CAAGEC}{CAAFTQ}_{t,q} - \frac{CAAGBM}{CAAFTQ}_{t,q} \tag{15}$$

In summary, we define the household, government and enterprise shares of total GFCF as

$$HHShGFCF_{t,q} = \frac{CAAGBM}{CAAFTQ}_{t,q} \tag{16}$$

[7]Of course, this assumes that the average of the quarterly shares equals the annual share which will not in general be true. However, the discrepancy is likely to be small.

[8]When estimating this, we find that the mean of the residual $\hat{e}_{t,q}$ is non-zero. Consequently, we estimate $\hat{e}_{t,q} = \alpha + \rho\hat{e}_{t,q-1}$by OLS without restricting $\alpha = 0$.

[9]Indeed, the revenue minus consumption share is negative prior to 2000, demonstrating the series are perhaps not mutually consistent.

$$GovtShGFCF_{t,q} = \frac{CAAGBM}{CAAFTQ}_{t,q} \tag{17}$$

$$EnterpriseShGFCF_{t,q} = 1 - \frac{CAAGEC}{CAAFTQ}_{t,q} - \frac{CAAGBM}{CAAFTQ}_{t,q} \tag{18}$$

To split enterprise investment into "SOEs", "private" and "other non-SOE", we use fixed assets investment (FAI). For all years from 1996 to 2013 (except 2010 for some reason) total annual FAI is the sum of the eight subcomponents in table 2. For 1995, we define

$$\begin{aligned} & (COMAD + COMAE + COMAH)_{t=1995} \\ = \; & COMA_{t=1995} - (COMAA_{t=1995} + COMAB_{t=1995} + \\ & COMAC_{t=1995} + COMAF_{t=1995} + COMAG_{t=1995}) \end{aligned} \tag{19}$$

and

$$COMAD_{t=1995} = (COMAD{+}COMAE{+}COMAH)_{t=1995} \frac{COMAD_{t=1996}}{(COMAD + COMAE + COMAH)_{t=1996}} \tag{20}$$

$$COMAE_{t=1995} = (COMAD{+}COMAE{+}COMAH)_{t=1995} \frac{COMAE_{t=1996}}{(COMAD + COMAE + COMAH)_{t=1996}} \tag{21}$$

$$COMAH_{t=1995} = (COMAD{+}COMAE{+}COMAH)_{t=1995} \frac{COMAH_{t=1996}}{(COMAD + COMAE + COMAH)_{t=1996}} \tag{22}$$

As shown in table 2, we assign each of the eight variables in that table to "SOE", "Private", and "Other Non-SOE".

$$FAIannSOE = COMAA + \frac{2}{3}COMAD \tag{23}$$

$$FAIann\Pr iv = COMAC + COMAH \tag{24}$$

$$FAIannOthNonSOE = COMAB + COMAE + COMAF + COMAG + \frac{1}{3}COMAD \tag{25}$$

$$FAIannTotal = FAIannSOE + FAIannPriv + FAIannOthNonSOE \tag{26}$$

The shares of annual FAI by sector are $ShareFAIannSOE = \frac{FAIannSOE}{FAIannTotal}$, $ShareFAIann\Pr iv = \frac{FAIannPriv}{FAIannTotal}$ and $ShareFAIannOthNonSOE = \frac{FAIannOthNonSOE}{FAIannTotal}$.

Staring in 2004, CEIC has monthly data on FAI for domestic enterprises that can be partitioned according to the variables in table 3. As with the annual variables, we assign each of the variables in table 3 to "SOE", "Private", and "Other Non-SOE".

$$FAIytdSOE = COBDLV + COBDLZ + COBDME + \frac{1}{2}COBDMB \tag{27}$$

$$FAIytdPriv = COBDMI + COBDMH + COBDMT \tag{28}$$

$$\begin{aligned} FAIytdOthNonSOE &= COBDLW + COBDLX + COBDMA + COBDMC \\ &\quad +COBDMF + COBDMG + COBDMJ \\ &\quad +COBDMO + .5 * COBDM \end{aligned} \tag{29}$$

$$FAIytdTotal = FAIytdSOE + FAIytdPriv + FAIytdOthNonSOE \tag{30}$$

Also, as with the annual shares, we define $ShareFAIytdSOE = \frac{FAIytdSOE}{FAIytdTotal}$, $ShareFAIytd\Pr iv = \frac{FAIytdriv}{FAIytdTotal}$ and $ShareFAIytdOthNonSOE = \frac{FAIytdOthNonSOE}{FAIytdTotal}$. For each of $FAIytdSOE$, $FAIytdPriv$, and $FAIytdOthNonSOE$ we undo the ytd conversion to get quarterly levels and seasonally adjust the resulting series. We call these seasonally adjusted series **FAIqSOE_SA**, **FAIqPriv_SA**, and **FAIqOthNon_SA** respectively.

The annual FAI shares have a level shift in 2006 which we eliminate as follows. Let $ShareFAIytdX$ and $ShareFAIannX$ denote the year-to-date monthly and annual FAI shares for sector X ("SOE", "Private", or "Other non-SOE"). We set $ShareFAIannXBrAdj_t = ShareFAIannX_t$ for $t \geq 2006$; $ShareFAIannXBrAdj_{t=2005} = ShareFAIannXBrAdj_{t=2006} + (ShareFAIytdX_{t=Dec2005} - ShareFAIytdX_{t=Dec2006})$; and $ShareFAIannXBrAdj_h = ShareFAIannX_h + (ShareFAIannXBrAdj_{t=2005} - ShareFAIannX_{t=2005})$ for $h < 2005$.

The original annual and break adjusted annual shares FAI shares are plotted in figure 14. The break-adjusted annual shares are multiplied by $FAIannTotal$ to get the break-adjusted levels $FAIannSOEBrAdj$, $FAIann\Pr ivBrAdj$, and $FAIannOthNonBrAdj$. These annual series, which cover 1995-2013, are each interpolated to the quarterly frequency using the minimal squared growth difference interpolation algorithm described in Appendix C. These interpolated series are padded on to the beginning of $FAIqSOE\_SA$, $FAIqPriv\_SA$, and $FAIqOthNon\_SA$ with the ratio splice backcasting method described in Appendix B (the splice point is 2004Q1). The spliced series are shown in figure 15 (the green dashed lines).

Finally, these spliced series are used interpolate $FAIannSOEBrAdj$, $FAIann\Pr ivBrAdj$, and $FAIannOthNonBrAdj$[10], which we name $FAISOEBrAdj\_Q$, $FAI\Pr ivBrAdj\_Q$, and $FAIOthNonBrAdj\_Q$. The growth rates of the annual series and the quarterly series are plotted in figure 16. Not surprisingly, the quarterly series is much more volatile starting in 2004 when the monthly year-to-date FAI data becomes available. The interpolated annual series are converted to the shares $ShFAISOEBrAdj\_Q$, $ShFAI\Pr ivBrAdj\_Q$, and $ShFAIOthNonBrAdj\_Q$ by dividing by the sum ($FAISOEBrAdj\_Q+FAI\Pr ivBrAdj\_Q+FAIOthNonBrAdj\_Q$). Referring back to equations (16) - (18), we define the following quarterly shares of GFCF:

-"Private" share of GFCF: $ShFAI\Pr ivBrAdj\_Q(1-HHShGFCF)$

-"Other Non-SOE" share of GFCF: $ShFAIOthNonBrAdj\_Q(1-HHShGFCF)$

-"SOE" share of GFCF: $FAISOEBrAdj\_Q(1-HHShGFCF)$

-"SOE excluding government" share of GFCF: $FAISOEBrAdj\_Q(1-HHShGFCF)-GovtShGFCF$

These quarterly shares, along with the governement and household sector shares are plotted in figure 17. These quarterly shares are all multiplied by quarterly nominal gross fixed capital formation [**NGFCF_Q_SA**] to get GFCF for each of the sector. These shares are multiplied by quarterly nominal gross fixed capital formation [**NGFCF_Q_SA**] to get GFCF for the five respective sectors. The quarterly growth rates of GFCF by sector are plotted in figure 18 and the HP filtered cycles are shown in figure 19[11]. The nominal GFCF series by sector are called **GovtGFCF_Q_SA**, **PrivNGFCF_Q_SA**, **OthNonSOENGFCF_Q_SA**, **SOEExGovtNGFCF_Q_SA**, and **HHNGFCF_Q_SA**.

II.4. **Industrial production.** The series we use – **IPFame_Q_SA** – is taken from the Federal Reserve's FAME database [ticker "JQI.M.CH"]. The analogous series from CEIC – "CBEOA; VAI: YoY Growth" – has two shortcomings. It has missing values in January or February in some recent years. Secondly, it often has large outliers in January or February. This is illustrated in figure 20. The FAME series has neither of these problems; it also has the advantage that it is a seasonally adjusted level. Note that the China's monthly

[10]First, the annual series are interpolated by the quarterly series for 2004-2013 with the adapted interpolation method of Fernandez (1981) described in Appendix A. The resulting quarterly series are backcasted with the original quarterly series using the ratio splice backcasting method described in Appendix D. Finally, these new quarterly series are used to interpolated the annual series (for 1995-2013) using proportional Denton interpolation.

[11]The annual flow of funds data does not define enterprise fixed capital formation by "SOE" and "private enterprise". We construct these by using the annual average of $BrAdjShareFAI_{t,q}^{SOE}$ and $1-BrAdjShareFAI_{t,q}^{SOE}$ multiplying by total enterprise fixed capital formation.

value added of industry is nominal[12]. Other authors appear to be using the nominal series as well. For example, the data appendix of Fernald, Malkin, and Spiegel (2013) uses the IP data from FAME (SA and NSA) as well as the aforementioned "CBEOA; VAI: YoY Growth" from CEIC which is nominal. The working paper by He, Leung, and Chong (2013) uses a series which they "Industrial Production Index: % Change"[13]. We are not aware of any series in CEIC with this label. It looks very similar to "CBEOA; VAI: YoY Growth" but not identical. It also has outliers in January and/or February of some years. Holz (2014) constructs both real and nominal monthly measures of industrial output for 1980-2012 though we do not use his data for our study.

II.5. **Disposable personal income and labor compensation.** The Flow of Funds database in CEIC has annual measures of (DPI) disposable personal income **DPI_A** [CEIC ticker CAAGFB], labor compensation **LaborComp_A** [CEIC ticker CAAFTY] and household income before tax **HHIncBefTax_A** [CEIC ticker CAAGEV] for 1992-2012. To interpolate **DPI_A** and **LaborComp_A**, we utilize the series in table 4. The first step is to construct a continuous time series $RuralCashInc_{t,q}^{PerCap}$ for "rural cash income per-capita". Figure 21 shows quarterly growth for the discontinued series CGEAA from 1996q1-2000q1 and the current series CGEAHA, where a continuous time series starts in 2000q4[14]. Splicing these series is tricky as figure 21 shows they have different seasonal patterns. Letting $d_{t,q}^{QQ} = \Delta\log(UndoYTD(CGEAA_{t,q}))$[15] and $q_{t,q}^{QQ} = \Delta\log(UndoYTD(CGEAHA_{t,q}))$, we calculate the average seasonal factors for $d_{t,q}^{QQ}$ over the 1996Q2 - 2000Q1 period and the average seasonal factors for $q_{t,q}^{QQ}$ over the 2001q2 - 2006q1 period[16]. Let d_h^{SAFact} denote the quarter h seasonal factor for $d_{t,q}^{QQ}$ (where $1 \leq h \leq 4$) and q_h^{SAFact} denote the quarter h seasonal factor for q_t^{QQ}. The change in the quarter h seasonal factor from 1996q2 - 2000q1 to 2001q2 - 2006q1 is $\tilde{q}_h^{SAFact} = q_h^{SAFact} - d_h^{SAFact}$. We solve for the unique value $RuralCashInc_{t=1996,q=1}^{PerCap}$[17] that satisfies

$$RuralCashInc_{t=1996,q=2}^{PerCap} = RuralCashInc_{t=1996,q=1}^{PerCap}\exp(d_{t=1996,q=2}^{QQ} + \tilde{q}_2^{SAFact}) \tag{31}$$

[12]From the CEIC documentation: Value added of industry is the difference between the total output value of production and the value of consumption and transfer of material products and service during production. **Value added of industry is calculated at current price.**

[13]This series is plotted in figure 6 of their paper.

[14]The CEIC series are ytd; the ytd transformation has been undone.

[15]$UndoYTD(x_{t,q})$ is defined as $UndoYTD(x_{t,1}) = x_{t,1}$ and $UndoYTD(x_{t,q}) = x_{t,q} - x_{t,q-1}$ when $q > 1$.

[16]The seasonal factors are the demeaned average growth rates by quarter; by construction the four seasonal factors for both d_t^{QQ} and q_t^{QQ} sum to 0.

[17]The soultion is found with the MatLab function fzero.

$$RuralCashInc^{PerCap}_{t=1996,q=3} = RuralCashInc^{PerCap}_{t=1996,q=2} \exp(d^{QQ}_{t=1996,q=3} + \tilde{q}^{SAFact}_{3}) \tag{32}$$

$$RuralCashInc^{PerCap}_{t=1996,q=4} = RuralCashInc^{PerCap}_{t=1996,q=3} \exp(d^{QQ}_{t=1996,q=4} + \tilde{q}^{SAFact}_{4}) \tag{33}$$

$$CGEAHA_{t=1996,q=4} = \sum_{q=1}^{4} RuralCashInc^{PerCap}_{t=1996,q} \tag{34}$$

Using the value $RuralCashInc^{PerCap}_{t=1996,q=4}$ we found in the previous step, we find the unique adjustment factor adj_{1997} that solves

$$RuralCashInc^{PerCap}_{t=1997,q=1} = RuralCashInc^{PerCap}_{t=1996,q=4} \exp(d^{QQ}_{t=1997,q=1} + \tilde{q}^{SAFact}_{1} + adj_{1997}) \tag{35}$$

$$RuralCashInc^{PerCap}_{t=1997,q=2} = RuralCashInc^{PerCap}_{t=1997,q=1} \exp(d^{QQ}_{t=1997,q=2} + \tilde{q}^{SAFact}_{2} + adj_{1997}) \tag{36}$$

$$RuralCashInc^{PerCap}_{t=1997,q=3} = RuralCashInc^{PerCap}_{t=1997,q=2} \exp(d^{QQ}_{t=1997,q=3} + \tilde{q}^{SAFact}_{3} + adj_{1997}) \tag{37}$$

$$RuralCashInc^{PerCap}_{t=1997,q=4} = RuralCashInc^{PerCap}_{t=1997,q=3} \exp(d^{QQ}_{t=1997,q=4} + \tilde{q}^{SAFact}_{4} + adj_{1997}) \tag{38}$$

$$CGEAHA_{t=1997,q=4} = \sum_{q=1}^{4} RuralCashInc^{PerCap}_{t=1997,q} \tag{39}$$

As shown in table 4, we already have $RuralCashInc^{PerCap}_{t=1998,q=1} = CGEAHA_{t=1998,q=1}$. We find the unique adjustment factor adj_{1998} that solves

$$RuralCashInc^{PerCap}_{t=1998,q=2} = RuralCashInc^{PerCap}_{t=1998,q=1} \exp(d^{QQ}_{t=1998,q=2} + \tilde{q}^{SAFact}_{2} + adj_{1998}) \tag{40}$$

$$RuralCashInc^{PerCap}_{t=1998,q=3} = RuralCashInc^{PerCap}_{t=1998,q=2} \exp(d^{QQ}_{t=1998,q=3} + \tilde{q}^{SAFact}_{3} + adj_{1998}) \tag{41}$$

$$RuralCashInc^{PerCap}_{t=1998,q=4} = RuralCashInc^{PerCap}_{t=1998,q=3} \exp(d^{QQ}_{t=1998,q=4} + \tilde{q}^{SAFact}_{4} + adj_{1998}) \tag{42}$$

$$CGEAHA_{t=1998,q=4} = \sum_{q=1}^{4} RuralCashInc_{t=1998,q}^{PerCap} \tag{43}$$

Similarly, setting $RuralCashInc_{t=1999,q=1}^{PerCap} = CGEAHA_{t=1999,q=1}$, we find the unique adjustment factor adj_{1999} that solves

$$RuralCashInc_{t=1999,q=2}^{PerCap} = RuralCashInc_{t=1999,q=1}^{PerCap} \exp(d_{t=1999,q=2}^{QQ} + \tilde{q}_2^{SAFact} + adj_{1999}) \tag{44}$$

$$RuralCashInc_{t=1999,q=3}^{PerCap} = RuralCashInc_{t=1999,q=2}^{PerCap} \exp(d_{t=1999,q=3}^{QQ} + \tilde{q}_3^{SAFact} + adj_{1999}) \tag{45}$$

$$RuralCashInc_{t=1999,q=4}^{PerCap} = RuralCashInc_{t=1999,q=3}^{PerCap} \exp(d_{t=1999,q=4}^{QQ} + \tilde{q}_4^{SAFact} + adj_{1999}) \tag{46}$$

$$CGEAHA_{t=1999,q=4} = \sum_{q=1}^{4} RuralCashInc_{t=1999,q}^{PerCap} \tag{47}$$

For 2000, we do not have good data for rural cash income and are forced to rely on the following crude interpolation. Setting $RuralCashInc_{t=2000,q=1}^{PerCap} = CGEAHA_{t=2000,q=1}$, we find the unique growth rate g_{2000} that solves

$$RuralCashInc_{t=2000,q=2}^{PerCap} = RuralCashInc_{t=2000,q=1}^{PerCap} \exp(q_2^{SAFact} + g_{2000}) \tag{48}$$

$$RuralCashInc_{t=2000,q=3}^{PerCap} = RuralCashInc_{t=2000,q=2}^{PerCap} \exp(q_3^{SAFact} + g_{2000}) \tag{49}$$

$$RuralCashInc_{t=2000,q=4}^{PerCap} = RuralCashInc_{t=2000,q=3}^{PerCap} \exp(q_4^{SAFact} + g_{2000}) \tag{50}$$

$$CGEAHA_{t=2000,q=4} = \sum_{q=1}^{4} RuralCashInc_{t=2000,q}^{PerCap} \tag{51}$$

We set $RuralCashInc_{t,q}^{PerCap}$ for $t \geq 2001$ by undoing the year-to-date operation on $CGEAHA$. This gives us our continuous time series time series $RuralCashInc_{t,q}^{PerCap}$, which we seasonally adjust along with $UrbanDPI_{t,q}^{PerCap}$. $UrbanDPI_{t,q}^{PerCap}$ is available starting in 2002Q1 and comes directly from undoing the year-to-date operation on CEIC variable CHAMBG. We backcast $UrbanDPI_{t,q}^{PerCap}$ by estimating a "reverse BVAR" with $y_{t,q} = [\log(UrbanDPI_{t,q}^{PerCap,SA}), \log(IP_{t,q}^{FAME,SA}), \log(CPI_{t,q}^{SA})]'$. The reverse BVAR takes the form

$$y_{t,q} = \sum_{h=1}^{3} A_h y_{t,q+h} + \varepsilon_{t,q} \tag{52}$$

We estimate (52) using data starting in 2002q2, backcast $\log(UrbanDPI_{t,q}^{PerCap,SA})$ for quarters between 1995q4 and 2001q4 using the conditional forecasting technique in Waggoner and Zha (1999), re-estimate (52) from 1995q4-present using the backcast augmented data, and repeat the conditional backcasting/BVAR estimation until the backcasts converge.

The next step is to convert $RuralCashInc_{t,q}^{PerCap,SA}$ and $UrbanDPI_{t,q}^{PerCap,SA}$ from per-capita to total income. We interpolate the log of the annual CEIC series "CGGA – Population" using a cubic spline and denote the quarterly series as $Pop_{t,q}$. The annual series measures the population at the end of each year while the interpolated quarterly series measures the population in the middle of each quarter. The annual rural share of the urban population is measured using the ratio of CEIC tickers $\frac{CGGE}{CGGE+CGGD}$ [see table 4]. The annual series has breaks as seen in figure 22. To remove these breaks, we set all values of the annual series $\frac{CGGE}{CGGE+CGGD}$ to missing values except for 1949, 1960, 1970, 1982, 1995, 2000, 2005, 2010 and 2014[18]. We then interpolate the resulting annual series with a cubic spline; the interpolated series $PopRuralShare_{t,q}$ is also shown in figure 22.

We set

-$PopRural_{t,q} = Pop_{t,q}$

-$PopRuralShare_{t,q}$

-$PopUrban_{t,q} = Pop_{t,q}(1 - PopRuralShare_{t,q})$

-$RuralCashInc_{t,q}^{SA} = RuralCashInc_{t,q}^{PerCap,SA} PopRural_{t,q}$

-$UrbanDPI_{t,q}^{SA} = UrbanDPI_{t,q}^{PerCap,SA} PopUrban_{t,q}$, and

-$RuralUrbanDPI_{t,q}^{SA} = RuralCashInc_{t,q}^{SA} + UrbanDPI_{t,q}^{SA}$.

We interpolate each of **DPI_A** and **LaborComp_A** with $RuralUrbanDPI_{t,q}^{SA}$ and the Fernandez (1981) method described in Appendix A to get **DPI_Q_SA** and **LaborComp_Q_SA**. Figure 23 shows $RuralUrbanDPI_{t,q}^{SA}$ along with annual and quarterly measures of labor compensation and DPI growth. We see that the annual growth rate of $RuralUrbanDPI_{t}^{SA}$ is highly correlated with both the annual growth rate of **DPI_A** ($r = 0.88$) and **LaborComp_A** ($r = 0.96$). This suggests that $RuralUrbanDPI_{t,q}^{SA}$ is a good interpolater. Figure 23 also shows the growth rates of the interpolated series **DPI_Q_SA** and **LaborComp_Q_SA**. In figure 23, **DPI_Q_SA** and **LaborComp_Q_SA** have been extrapolated through 2013Q4 by $RuralUrbanDPI_{t,q}^{SA}$ using a BVAR.

[18]These dates are roughly chosen to correspond to Chinese population census dates.

II.6. **Financial indicators. Central bank policy rate (SpliceRepo1Day_Q):** The CEIC series "CMPAL – National Interbank Bond Repurchase: WA Rate: NIBFC: 1 Day" is available monthly since March 2003; it is the primary series used for **SpliceRepo1Day_M**. The CEIC series "CMOAA – National Interbank Offered Rate: Weighted Avg: NIBFC: Overnight" is spliced with **SpliceRepo1Day_M** back to July 1996 using the additive splicing method described in Appendix B[19]. Finally, the CEIC series "CMOAB – National Interbank Offered Rate: Weighted Avg: NIBFC: 7 Days" is spliced with **SpliceRepo1Day_M** back to January 1996 using additive splicing. **SpliceRepo1Day_Q** is the quarterly averages of **SpliceRepo1Day_M**. **SpliceRepo1Day_M** and its component series is plotted in figure 24.

Monetary aggregates M0 and M2 (M0_Q_SA and **M2_Q_SA):** Table 5 provides the series used (the steps for constucting M0 and M2 are identical). The target series $yoy^*_{t,m}$ is initialized with the published series. We calculate the series $yoy^{proxy}_{t,m} = 100(\frac{Q_{t,m}}{Q_{t-1,m}} - 1)$ which is available at the end of each quarter. We linearly interpolate $yoy^{proxy}_{t,m}$ to fill in the missing values [calling the new series $y\tilde{o}y^{proxy}_{t,m}$]. Finally, whenever $yoy^*_{t,m}$ is missing and $y\tilde{o}y^{proxy}_{t,m}$ is not, we replace the missing value for $yoy^*_{t,m}$ with $y\tilde{o}y^{proxy}_{t,m}$. Suppose time $(YEAR = t_1, MONTH = m_1)$ has the most recent observation of $M_{t,m}$. For $m_1 - 11 \leq m \leq m_1$, we set $M^*_{t,m} = M_{t,m}$. For $m < m_1 - 11$ we use backward recursion to define $M^*_{t_1,m} = \frac{100M^*_{t_1,m+12}}{100+yoy^*_{t_1,m+12}}$. The monthly series are aggregated to quarterly levels and seasonally adjusted. The 12-month percent changes of the NSA series and the quartely log changes of the SA series are plotted in figure 25. The 12-month percent change of M0 has outliers around January and February of some years, but these are mostly eliminated after quarterly averaging and seasonal adjustment.

Reserves and reserve ratios (**PBOCReserves_AvgQ_SA**, **RequiredReserveRatio_Q**, **ARR_Q_SA**, and **ERR_Q_SA**): The building block for aggregate reserves is the monthly CEIC series "CKDP – Monetary Authority: Liab: Reserve Money (RM)" that is available since December 1999. The quarterly average of this series (**CKDP_Q**) is used. An end-of-quarter series "CKDA – Quarterly Monetary Authority: Liab: Reserve Money (RM)" is available since March 1993. The log of this series is linearly interpolated to the monthly frequency, then exponentiated, then quarterly averaged to get **CKDA_Q**. **PBOCReserves_AvgQ** is set to **CKDP_Q** backcasted by **CKDA_Q** with the ratio splice backcasting method described in Appendix B. The seasonally adjusted series is **PBOCReserves_AvgQ_SA**. **RequiredReserveRatio_Q** is the quarterly average of the monthly CEIC series "CMAAAA – CN: Required Reserve Ratio". The total reserve ratio series

[19]CMOAA does have occasional missing values prior to March 2000; prior to the splicing we replace these missing values with an average of the values in the prior and subsequent month.

is defined as **ARR_Q_SA**=100(**PBOCReserves_AvgQ_SA - M0_Q_SA**)/(**M2_Q_SA - M0_Q_SA**). The excess reserve ratio is defined as **ERR_Q_SA = ARR_Q_SA - RequiredReserveRatio_Q**. The reserve ratios are plotted in figure 26.

II.7. **Price indexes.** Table 6 lists the components used to construct the monthly NSA CPI and PPI. Let (t_0, m_0) be the first date $pp100_{t,m}$ is available. We set $\tilde{p}_{t_0,m_0-1} = 100$ and recursively define $\tilde{p}_{t_0,m} = \frac{pp100_{t_0,m}}{100}\tilde{p}_{t_0,m-1}$ for $m \geq m_0$. Let (t_1, m_1) be the last date $\tilde{p}_{t,m}$ is available. We set $p^*_{t_1,m} = \tilde{p}_{t_1,m}$ for $m_1 - 11 \leq m \leq m_1$ and use backward recursion to define

$$p^*_{t_1,m} = \frac{100 p^*_{t_1,m+12}}{py100_{t_1,m+12}} \tag{53}$$

for $m < m_1 - 11$. This gives us the monthly NSA levels **CPI_M** and **PPI_M**. For the retail price index, CEIC has the series "CIXA – CN: Retail Price Index, Prev Year=100" but does not appear to have a level or a "previous month=100 series". Since January 1987 is the first observation for CIXA, we set **RetailPrice_M** = **CPI_M** for each of the 12 months in 1986 and recursively define **RetailPrice_M**(t) = (**RetailPrice_M**(t-12)CIXA(t))/100 for t $\geq$ Jan1987. Quarterly averages of the NSA monthly CPI, PPI and retail price series are seasonally adjusted. The data are plotted in figure 27.

The price of fixed assets investment (FAI) – **SplicePriceInvest_A** – uses "CITA – Fixed Assets Investment Price Index: Overall (PY = 100)" for 1989 to 2014 [by cumulating the inflation rate CITA]. For 1990-2013, the (log) inflation rate for FAI investment prices is regressed on the annual log inflation rates for the CPI, PPI and GDP Deflator. The regression coefficients are in table 7. PPI inflation is the only term with a coefficient estimate that is statistically significant. Nonetheless, if GDP deflator inflation is approximately a linear combination of consumption price inflation and investment price inflation, then the negative coefficient on the CPI and the positive coefficient on the GDP deflator is what we would expect. The fitted values of the regression are cumulated backward to fill in **SplicePriceInvest_A** from 1978 to 1988.

Table 15 of Holz (2013b) has data on the inflation rate for the implicit price deflator for gross fixed capital formation (GFCF) for 1979-2004. As explained in the appendix to the paper, Holz (2013a), the data originally come from NBS (2007). Apparently, the GFCF price deflator is no longer released by the NBS. However, over the common 1990-2004 sample period, the log first difference of the GFCF price deflator [**IndexGFCFPriceHolz**] has a correlation of 0.994 with the log first difference of **PriceInvest_A**. Therefore we regress $\Delta\log$(**IndexGFCFPriceHolz**) on a constant and $\Delta\log$(**PriceInvest_A**) and use the regression to extrapolate **IndexGFCFPriceHolz** to 2013. The regression results are in table 8. The resulting series is called **GFCFPriceHolzSplice_A**.

The upper panel of figure 28 plots the annual inflation rates for four of the price deflators and the log prices of **PPI_A**, **SplicePriceInvest_A**, **GFCFPriceHolzSplice_A** relative to the log of **CPI_A**. We see that, puzzlingly, the inflation rates for GFCF and FAI lead the inflation rates for the PPI and CPI over the 1992-1995. This leads to the sawtooth pattern for the relative price of investment for GFCF and FAI shown in the lower panel of figure 28. There is only a modest spike in the relative price of the PPI over the same period. This suggests that sawtooths for the relative prices in GFCF and FAI from 1990-1995 could be spurious.

II.8. **Annual consumer price indexes for durables, nondurables and services.** Our measures of durable and nondurable CPIs are constructed as Tornqvist aggregates of item-level CPIs. The expenditure weights are derived from annual data on per-capita consumption expenditures of urban and rural households available in the *China Statistical Yearbook (CSY)*[20]. The CEIC tickers for these items are in table 9. Unavailable rural expenditures for disaggregated items are proxied by their urban counterparts. The per-capita series are converted to total expenditures using the rural and urban population measures defined in the earlier section on labor income. Total consumption by item is defined as the sum of urban and rural consumption. Table 10 provides the CEIC tickers for nondurable goods CPI items and the expenditure weights from table 9. Table 11 does the same for durable goods.

Since there isn't a perfect concordance between consumption expenditures and CPI items, some assumptions must be made. For example, since the CPI item "Educ.: Teaching Mat. & Refer. Book" is a durable consumption good according to the U.S. classification, it is assigned as a durable good for China as well. Since it is probably only a small portion of total education expenditures, we use $\frac{1}{20}$th of total educational expenditures as its weight. The CPI tickers in tables 10 and 11 are all "pervious year = 100". We convert these to log changes $\Delta \log p_{i,t}$. The inflation rate for nondurable goods in year t is:

$$\Delta \log p_t^{ND} = \sum_{i \in NonDur} \{\frac{1}{2}[\frac{shND_iC_{i,t}}{\sum\limits_{j \in NonDur} shND_jC_{j,t}} + \frac{shND_iC_{i,t-1}}{\sum\limits_{j \in NonDur} shND_jC_{j,t-1}}]\} \Delta \log p_{i,t} \qquad (54)$$

where the $shND_iC_{i,t}$ terms are weighted consumption levels shown in columns 3 and 4 of table 10. The use of different weights pre-2001 and since 2001 is needed as some of the CPI data starts in 2001. Durable goods CPI inflation is defined analagously:

[20]For example, see tables 11-15 and 11-24 of the 2013 CSY.

$$\Delta \log p_t^{Dur} = \sum_{i \in Durables} \{\frac{1}{2}[\frac{shDur_i C_{i,t}}{\sum_{j \in Durables} shDur_j C_{j,t}} + \frac{shDur_i C_{i,t-1}}{\sum_{j \in Durables} shDur_j C_{j,t-1}}]\} \Delta \log p_{i,t} \quad (55)$$

The distinct items in table 9 as well as "residence" consumption[21] partition the total consumption basket. The total Tornqvist weights for consumer nondurables and consumer durables are:

$$wt_t^{ND} = \sum_{i \in NonDur} \{\frac{1}{2}[\frac{shND_i C_{i,t}}{\sum_{j \in ConsumptionBasket} C_{j,t}} + \frac{shND_i C_{i,t-1}}{\sum_{j \in ConsumptionBasket} C_{j,t-1}}]\} \quad (56)$$

$$wt_t^{Dur} = \sum_{i \in NonDur} \{\frac{1}{2}[\frac{shDur_i C_{i,t}}{\sum_{j \in ConsumptionBasket} C_{j,t}} + \frac{shDur_i C_{i,t-1}}{\sum_{j \in ConsumptionBasket} C_{j,t-1}}]\} \quad (57)$$

The Tornqvist weight for consumption services is

$$wt_t^{Svc} = 1 - wt_t^{ND} - wt_t^{Dur} \quad (58)$$

The annual inflation rate for CPI sevices is available in CEIC as the ticker CIKQ ["CN: CPI: Service: Prev Year=100] for all years since 1986 except for 2001-2003 and 2007-2008. We convert it to the log inflation rate $\Delta \log p_t^{Svc}$. Total CPI inflation $\Delta \log p_t^{Tot}$ is derived from the CEIC ticker CIKG ["Consumer Price Index: PY=100"]. For the years 2001-2003 and 2007-2008 where $\Delta \log p_t^{Svc}$ is missing, we define it implictly from:

$$\Delta \log p_t^{Svc} = \frac{\Delta \log p_t^{Tot} - wt_t^{ND} \Delta \log p_t^{ND} - wt_t^{Dur} \Delta \log p_t^{Dur}}{wt_t^{Svc}} \quad (59)$$

For the years where $\Delta \log p_t^{Svc}$ is available from CEIC, we adjust durable goods and non-durable goods CPI inflation by defining the adjustment factor

$$\pi AdjFact_t^{Goods} = \frac{\Delta \log p_t^{Tot} - wt_t^{ND} \Delta \log p_t^{ND} - wt_t^{Dur} \Delta \log p_t^{Dur} - wt_t^{Svc} \Delta \log p_t^{Svc}}{1 - wt_t^{Svc}} \quad (60)$$

and defining adjusted durable goods and nondurable goods CPI inflation as

$$\Delta \log \tilde{p}_t^{ND} = \Delta \log p_t^{ND} + \pi AdjFact_t^{Goods} \quad (61)$$

$$\Delta \log \tilde{p}_t^{Dur} = \Delta \log p_t^{Dur} + \pi AdjFact_t^{Goods} \quad (62)$$

For the years 2001-2003 and 2007-2008 we set $\pi AdjFact_t^{Goods} = 0$ and define the price levels **AdjCPINonDur_A**, **AdjCPIDur_A**, **AdjCPISVC_A**, **CPINonDur_A**, and **CPIDur_A** by taking the cumulative products of the exponentials of $\Delta \log \tilde{p}_t^{ND}$, $\Delta \log \tilde{p}_t^{Dur}$, $\Delta \log p_t^{Svc}$,

[21]CEIC ticker CHABDR is urban per-capita residence spending while CHABLS is rural.

$\Delta \log p_t^{ND}$, and $\Delta \log p_t^{Dur}$. These are plotted in figure 29. The upper left panel includes a second measure of consumer durables inflation **PPIConsDur_A** which is the PPI measure derived from CEIC ticker CIACYC. We see that for both durables and nondurables, the diffference between the "original" and "adjusted" inflation rates are minimal since 2003. This gives us some confidence that we are constructing these aggregates in a sensible way.

II.9. **Quarterly consumer price indexes for durables, nondurables and services.** Our first step is to define a set of item-level quarterly seasonally adjusted CPI price indexes that we can use to interpolate the series in tables 10 and 11. The CEIC tickers used to construct interpolaters are shown in table 12. These NSA monthly CPI and retail price index series often must be spliced together. The operator "/" uses the standard splice method while the operator "#" exponentiates after additive splicing of the logarithms [both of these are described in Appendix B]. For example, for "Health Care Appliance, Articles & Medical Inst", the monthly CPI ticker "CIAHUN" starts in January 2005. The log of this series is additively spliced with the log of the CEIC ticker "CIXF – Retail Price Index: Medicines, Medical & Health Care Articles (PY=100)" that begins in January 2001. The exponential of the resulting series is then "standard spliced" to the discontinued CPI series "CIAK – Medicines and Medical Care (PY = 100)".

All of the spliced NSA series are converted to quarterly price levels, seasonally adjusted and used to interpolate the annual CPI series listed in tables 10 and 11 as well as the additional series defined in table 13. Fernandez (1981)'s adapted interpolation routine described in Appendix A is used. In several cases the quarterly series starts before the annual series[22]; in such cases the interpolated series is backcasted with the quarterly series using the ratio splice backcast described in Appendix B. Tables 14 and 15 list the quarterly CPIs used to create Tornqvist indexes for durable and nondurable goods prices. The expenditure categories used to create the Tornqvist weights are defined in table 9 above. The shares are created at the annual frequency and then linearly interpolated to the quarterly frequency[23]. These Tornqvist aggregates – **calcDurCPI_Q** and **calcNondurCPI_Q** – are used to interpolate the price indices **AdjCPINonDur_A** and **AdjCPIDur_A** created earlier. The interpolated series **AdjCPIDur_Q** [$\tilde{p}_{t,q}^{Dur}$] and **AdjCPINonDur_Q** [$\tilde{p}_{t,q}^{ND}$] are used with the overall quarterly CPI index **CPI_Q_SA** [$p_{t,q}^{Tot}$] to define the seasonally adjusted quarterly services CPI interpolater **calcAdjCPISvc_Q** [$p_{t,q}^{Svc}$] as:

[22] "Fuel & Parts", "Personal Article & Services", and "Communication facilities".

[23] The interpolation is done by assigning year y to time $y + 0.5$ and quarter $yQTRq$ to time $y + \frac{2q-1}{8}$. Shares at tbe end (beginning) of the time series are carried over for later (earlier) periods.

$$\Delta \log p_{t,q}^{Svc} = \frac{\Delta \log p_{t,q}^{Tot} - wt_{t,q}^{ND} \Delta \log \tilde{p}_{t,q}^{ND} - wt_{t,q}^{Dur} \Delta \log \tilde{p}_{t}^{Dur}}{1 - wt_{t,q}^{ND} - wt_{t,q}^{Dur}} \tag{63}$$

where the weights are linear interpolations of (56) and (57). This series is used to interpolate the annual series **AdjCPISVC_A** with the Fernandez (1981) interpolation routine defined in Appendix A; the interpolated series is **AdjCPISVC_Q**. The quarterly inflation rates along with their annual counterparts are plotted in figure 30. We see that quarterly movements in the interpolated series very close match quarterly movements in the interpolaters.

II.10. **Loans.** For short-term, medium and long-term, and total financial institution loans outstanding, we use monthly CEIC data (tickers CKAHLA, CKAHLB and CKSAC, respectively). Prior to April 1999 we use monthly data from WIND for these three categories (no splicing is used). After aggregating the loans to the quarterly frequency and seasonally adjusting them, the three loan series are called **BankLoansST_Q_SA, BankLoansMLT_Q_SA** and **BankLoans_Q_SA** The data are plotted in figure 31. Annual data on the flow of loans to non financial enterprises for these three size categories – **NFELoanFOF_A**, **NFESTLoanFOF_A**, and **NFEMLTLoanFOF_A** – come from CEIC tickers CAACBD, CAAGNJ and CAAGNK, respectively. As seen in figure 32, **NFEMLTLoanFOF_A** is almost identical to December values of the monthly year-to-date CEIC series "CKATXZ – CN: Loan: New Increased: ytd: Non FI & Other: Medium & Long Term"; therefore we extrapolate the 2013 value of it with the December 2013 value of CKATXZ. Also as seen in figure 32, for 2011 and 2012 **NFESTLoanFOF_A** is quite close to the sum of the December value of the CEIC series "CKATXX – CN: Loan: New Increased: ytd: Non FI & Other: Short Term" and "CKATXY – CN: Loan: New Increased: ytd: Non FI & Other: Bill Financing"[24]. Therefore, we set the 2013 value of **NFESTLoanFOF_A** to the sum of the December 2013 values of CKATXX and CKATXY.

CEIC has end-of-quarter data on both medium+long-term – **LoanUseNonfinEntMLT_Q** (CEIC ticker CKANLT) – and short-term+bill financing – **LoanUseNonfinEntSTBF_Q** (CEIC ticker CKANLQ) – non-financial loans outstanding since 2007. Quarterly measures of "new increased" non-financial institution loans for both the medium+long-term – **NewIncLoanNonFIOthMLT_Q** – and short-term+bill financing – **NewIncLoanNonFIOthSTBF_Q** – categories are derived from monthly ytd measures from the CEIC tickers CKATXZ, CKATXX and CKATXY.

Figure 33 shows that the first difference of **LoanUseNonfinEntMLT_Q** is highly correlated with **NewIncLoanNonFIOthMLT_Q** while the first difference of **LoanUseNonfinEntSTBF_Q** is highly correlated with **NewIncLoanNonFIOthSTBF_Q**. Therefore,

[24]The correspondence is not as tight for 2007-2010.

we backcast the levels from 2006q4 to 2004q4 by cumulating the flows **NewIncLoanNonFIOthMLT_Q** and **NewIncLoanNonFIOthSTBF_Q** backwards[25]. We seasonally adjust the backcasted loan-use level series that start in 2004Q4 and then backcast the resulting series to 1994Q1 with the ratio splice method described in Appendix B using the series **BankLoansST_Q_SA** and **BankLoansMLT_Q_SA**. The resulting series are called **LoanUseNonfinEntSTBF_QbackSA** and **LoanUseNonfinEntMLT_QbackSA**. The 1994q4 values of these two series are used to initialize the 1994 levels of **LevNFESTLoanFOF_A** of **LevNFEMLTLoanFOF_A.** These levels, going forwards from 1994 to 2013 are filled in by accumulating the flows **NFESTLoanFOF_A** with **NFEMLTLoanFOF_A** using the recursion $Lev_{t+1} = Lev_t + Flow_{t+1}$. These levels **LevNFESTLoanFOF_A** of **LevNFEMLTLoanFOF_A**, are interpolated to the quarterly frequency by **LoanUseNonfinEntSTBF_QbackSA** and **LoanUseNonfinEntMLT_QbackSA**, respectively, using proportional first difference Denton interpolation[26]. The first differences of these interpolated series are named **NFESTLoanFOF_Q_SA** and **NFEMLTLoanFOF_Q_SA**, and they are the quarterly versions of **NFESTLoanFOF_A** and **NFEMLTLoanFOF_A**. In figure 34, we see that these series closely match the first differences of **LoanUseNonfinEntSTBF_QbackSA** and **LoanUseNonfinEntMLT_QbackSA**.

III. Implications

To be completed.

IV. Conclusion

To be completed.

[25] I.e., we do backwards recursion with the equation $Lev_{t,q-1} = Lev_{t,q} - Flow_{t,q}$.

[26] The function call in Matlab is essentially Denton_uniProp(YAnn,XQtr,2,1,4). This is modified slightly to handle the missing values that the original function does not allow.

Appendix A. Adapted Fernandez (1981) interpolation

Suppose we have an annual time series Y_t, where $1 \le t \le T$, and a related quarterly time series $X_{t,q}$, whose log difference, we believe, is a strong candidate interpolater series for $\Delta \log Y_t$. I.e., we believe the following model:

$$\Delta \log Y_{t,q} = [1, \Delta \log X_{t,q}]\boldsymbol{\beta} + \varepsilon_{t,q} \tag{A1}$$

where $Y_{t,q}$ is unobserved and, crucially, we assume $\varepsilon_{t,q}$ is iid $N(0, \sigma_\varepsilon^2)$. This distinguishes it from the weaker assumption for the Chow-Lin (1971) model that $\varepsilon_{t,q}$ is an AR(1) process. With the Fernandez (1981) problem, the interpolated series $\Delta \log \hat{Y}_{t,q}$ and the regression coefficients are the solution to the following problem:

$$\{\{\{\Delta \log \hat{Y}_{t,q}\}_{q=1}^4\}_{t=1}^T, \hat{\boldsymbol{\beta}}\} = \arg \min_{\{\{\Delta \log \hat{Y}_{t,q}\}_{q=1}^4\}_{t=1}^T, \boldsymbol{\beta}} \sum_{t=1}^T \sum_{q=1}^4 \{\Delta \log Y_{t,q} - [1, \Delta \log X_{t,q}]\boldsymbol{\beta}\}^2 \tag{A2}$$

subject to the constraints

$$\Delta \log Y_1 = \frac{1}{10}[4\Delta \log Y_{1,1} + 3\Delta \log Y_{1,2} + 2\Delta \log Y_{1,3} + \Delta \log Y_{1,4}] \tag{A3}$$

and

$$\begin{aligned} \Delta \log Y_t \quad = \quad & \frac{1}{16}[\Delta \log Y_{t-1,2} + 2\Delta \log Y_{t-1,3} + 3\Delta \log Y_{t-1,4} + 4\Delta \log Y_{t,1} \\ & +3\Delta \log Y_{t,2} + 2\Delta \log Y_{t,3} + \Delta \log Y_{t,4}] \end{aligned} \tag{A4}$$

for $2 \le t \le T$. Constraint (A4) is motivated by the approximation:

$$\begin{aligned} & \Delta \log \frac{Z_{t,1} + Z_{t,2} + Z_{t,3} + Z_{t,4}}{Z_{t-1,1} + Z_{t-1,2} + Z_{t-1,3} + Z_{t-1,4}} \\ \approx \quad & \frac{1}{16}[\Delta \log Z_{t-1,2} + 2\Delta \log Z_{t-1,3} + 3\Delta \log Z_{t-1,4} + \\ & 4\Delta \log Z_{t,1} + 3\Delta \log Z_{t,2} + 2\Delta \log Z_{t,3} + \Delta \log Z_{t,4}] \end{aligned} \tag{A5}$$

The solution to (A2) is derived by the following sequence of equations:

$$\mathbf{p}_1 = \frac{1}{16}[0, 1, 2, 3] \tag{A6}$$

$$\mathbf{p}_2 = \frac{1}{16}[4, 3, 2, 1] \tag{A7}$$

$$\mathbf{B} = \{\frac{1}{10}\begin{bmatrix} 4 & 3 & 2 & 1 & \mathbf{0}_{1x4(T-1)} \\ \mathbf{0}_{(T-1)x1} & \mathbf{0}_{(T-1)x1} & \mathbf{0}_{(T-1)x1} & \mathbf{0}_{(T-1)x1} & \mathbf{0}_{(T-1)x4(T-1)} \end{bmatrix} + \\ (\begin{bmatrix} \mathbf{0}_{1x(T-1)} & 0 \\ \mathbf{I}_{T-1} & \mathbf{0}_{(T-1)x1} \end{bmatrix} \otimes \mathbf{p}_1) + (\begin{bmatrix} 0 & \mathbf{0}_{1x(T-1)} \\ \mathbf{0}_{(T-1)x1} & \mathbf{I}_{T-1} \end{bmatrix} \otimes \mathbf{p}_2)\}' \tag{A8}$$

$$\mathbf{D} = \mathbf{I}_{4T} + \begin{bmatrix} \mathbf{0}_{1x(4T-1)} & 0 \\ -\mathbf{I}_{4T-1} & \mathbf{0}_{(4T-1)x1} \end{bmatrix} \tag{A9}$$

$$\mathbf{Z} = \begin{bmatrix} 1 & \Delta \log X_{1,1} \\ 2 & \Delta \log X_{1,2} \\ . & . \\ . & . \\ 4T-1 & \Delta \log X_{T,3} \\ 4T & \Delta \log X_{T,4} \end{bmatrix} \tag{A10}$$

$$\hat{\boldsymbol{\beta}} = [(\mathbf{Z}'\mathbf{B})(\mathbf{B}'(\mathbf{D}'\mathbf{D})^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{Z}]^{-1}\mathbf{Z}'\mathbf{B}(\mathbf{B}'(\mathbf{D}'\mathbf{D})^{-1}\mathbf{B})^{-1} \begin{bmatrix} \Delta \log Y_1 \\ \Delta \log Y_2 \\ . \\ . \\ \Delta \log Y_{T-1} \\ \Delta \log Y_T \end{bmatrix} \tag{A11}$$

$$\begin{bmatrix} \Delta \log \hat{Y}_{1,1} \\ \Delta \log \hat{Y}_{1,2} \\ . \\ . \\ \Delta \log \hat{Y}_{T,3} \\ \Delta \log \hat{Y}_{T,4} \end{bmatrix} = \mathbf{Z}\hat{\boldsymbol{\beta}} + (\mathbf{D}'\mathbf{D})^{-1}\mathbf{B}(\mathbf{B}'(\mathbf{D}'\mathbf{D})^{-1}\mathbf{B})^{-1}\{\begin{bmatrix} \Delta \log Y_1 \\ \Delta \log Y_2 \\ . \\ . \\ \Delta \log Y_{T-1} \\ \Delta \log Y_T \end{bmatrix} - \mathbf{B}'\mathbf{Z}\hat{\boldsymbol{\beta}}\} \tag{A12}$$

The matrix $\mathbf{B}$ imposes the constraints (A3) and (A4), while $\mathbf{D}$ is a differencing matrix. Once we have the estimated values $\{\{\Delta \log \hat{Y}_{t,q}\}_{q=1}^{4}\}_{t=1}^{T}$, we solve for the level by setting $\hat{Y}_{1,1} = \exp(\frac{11}{8}\log(Y_1) - \frac{3}{8}\log(Y_2))$and using the recursion

$$\hat{Y}_{t,q} = \hat{Y}_{t,q-1}\exp(\frac{\Delta \log \hat{Y}_{t,q}}{4}) \tag{A13}$$

Since equation (A5) is only an approximation, $\frac{1}{4}\sum_{q=1}^{4}\hat{Y}_{t,q} = Y_t$, will only hold approximately. To make sure it holds exactly, we use proportional Denton interpolation to interpolate Y_t with $\hat{Y}_{t,q}$. The MatLab programs are available at

http://www.mathworks.com/matlabcentral/fileexchange/24438-temporal-disaggregation-library

and the function call is denton_uni_prop(Ylf,Xhf,2,1,4) where Ylf is the low frequency variable Y_t and Xhf is the high frequency variable $\hat{Y}_{t,q}$. The interpolated value $\hat{Y}^*_{t,q}$ is multiplied by $\frac{1}{4}$ so that the sum of the quarterly values equals the annual total.[27] $\frac{1}{4}\hat{Y}^*_{t,q}$ is the time series returned in the program[28]

An obvious question is, why don't we simply use standard Chow-Lin (1971) interpolation. The Chow-Lin set-up would assume

$$Y_{t,q} = [1, X_{t,q}]\boldsymbol{\beta}_q + u_{t,q}$$

$$u_{t,q} = \rho_q u_{t,q-1} + e_{t,q}$$

with $e_{t,q}$ iid $N(0, \sigma_e^2)$. However, the level of GDP grows exponentially, so it is not plausible that $e_{t,q}$ has a constant variance. The variance should be larger at the end of the sample then the beginning. One could of course adapt Chow-Lin (1971) to handle data in log differences imposing constraints like (A3) and (A4).

Appendix B. Ratio splice, additive splice, and standard splice backcasting

Suppose we have a time series Y_t that is either of the monthly, quarterly, or annual frequency. Time $t+1$ is one period after time t. Suppose the first non-missing value for Y_t is time T_0 and we have a very closely related time series X_t that has the same frequency as Y_t but has non-missing values both in period T_0 and some periods prior to T_0. Then, the ratio spliced series Y_t^* is created by setting $Y_t^* = Y_t$ for $t \geq T_0$ and $Y_t^* = Y_{T_0}\frac{X_t}{X_{T_0}}$ for $t < T_0$. With additive splice backcasting, Y_t^* is created by setting $Y_t^* = Y_t$ for $t \geq T_0$ and $Y_t^* = X_t + (Y_{T_0} - X_{T_0})$ for $t < T_0$. With standard splice backcasting, Y_t^* is created by setting $Y_t^* = Y_t$ for $t \geq T_0$ and $Y_t^* = X_t$ for $t < T_0$.

---

[27] Unlike the Chinese NIPAs, in the U.S. NIPAs, the average of the quarterly values equals the annual total.

[28] For some series, the first non-missing quarter of the quarterly interpolater is the first quarter of a year T_0 where the growth rate of the annual series to be interpolated is available. When we want to interpolate the quarterly values for that year, we replace the missing value $\Delta \log X_{T_0,1}$ with $\frac{1}{4}(\sum_{q=2}^{4}\Delta \log X_{T_0,1} + \Delta \log X_{T_0+1,1})$.

Appendix C. Minimal squared growth difference interpolation

Suppose we have an annual time series Y_t, where $1 \leq t \leq T$ and we want to solve the following minimization problem

$$\{\{\{\Delta \log \hat{Y}_{t,q}\}_{q=1}^4\}_{t=1}^T\} = \arg \min_{\{\{\Delta \log Y_{t,q}\}_{q=1}^4\}_{t=1}^T} \sum_{t=1}^T \sum_{q=1}^4 \{\Delta \log Y_{t,q} - \Delta \log Y_{t,q-1}\}^2 \tag{A14}$$

subject to the constraints $(1 \leq t \leq T)$:

$$Y_t = 4 \prod_{q=1}^4 Y_{t,q}^{0.25} \tag{A15}$$

(A15) can be rewritten as

$$\log Y_t = \log(4) + \sum_{q=1}^4 0.25 \log \hat{Y}_{t,q} \tag{A16}$$

Let $\mathbf{y}^q = [\log(Y_{1,1}), \log(Y_{1,2}), \log(Y_{1,3}), ..., \log(Y_{T,3}), \log(Y_{T,4})]'$ and $\mathbf{y}^a = [\log(Y_1), \log(Y_2), ..., \log(Y_T)]'$. Let $\mathbf{D} = [d_{i,j}]$ be the $\{4T-1\}$ by $4T$ difference matrix with $d_{i,i}= -1$ and $d_{i,i+1}=1$ $(1 \leq i \leq 4T-1)$ and 0 for all other entries. Let $\mathbf{A} = \mathbf{I}_T \otimes [\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}]$. Then $\hat{\mathbf{y}}^q$ is the solution to the quadratic programming problem

$$\hat{\mathbf{y}}^q = \arg \min_{\mathbf{y}^q} \mathbf{y}^{q\prime} \mathbf{D}' \mathbf{D} \mathbf{y}^q$$

subject to

$$\mathbf{A}\mathbf{y}^q = \mathbf{y}^a - \log(4)\mathbf{1}_{Tx1}$$

This is easily solved with the Matlab function quadprog.m. Exponentiating the elements of $\hat{\mathbf{y}}^q$ gives the solution $\hat{\mathbf{Y}}^q$. Finally, as we want the constraint

$$Y_t = \sum_{q=1}^4 Y_{t,q} \tag{A17}$$

to hold rather the geometric mean constraint (A15), we use proportional Denton interpolation to interpolate Y_t with $\hat{Y}_{t,q}$.

Appendix D. Combining a monthly "previous year =100" series with a monthly "previous month =100" series

Let $Y_{t,m}$ denote a PY=100 series and $X_{t,m}$ denote a PM=100 series. Suppose the last date where both are available is month M_1 of year T_1. We set $Z_{T_1,M_1} = 1$ and, for $k = 0, 1, 2, ...$ recursively define

$$AdjFactYOY_{T_1,M_1-12k} = \frac{1}{12}(\log(\frac{Y_{T_1,M_1-12k}}{100}) - \sum_{h=0}^{11} \log(\frac{X_{T_1,M_1-12k-h}}{100})) \tag{A18}$$

$$Z_{T_1,M_1-12k-h} = Z_{T_1,M_1-12k-h+1} \exp(\log(\frac{X_{T_1,M_1-12k-h+1}}{100}) + AdjFactYOY_{T_1,M_1-12k}) \tag{A19}$$

where h runs from 1 to 12 in (A19). We continue these recursions until we first set Z_{T_0,M_0} to a missing value. At this point, it will often be the case that Y_{T_0,M_0+11} is non-missing as the PY=100 is available before the PM=100 series. In these cases we use backwards recursion to define

$$Z_{T_0,M_0-h} = \frac{100 Z_{T_0,M_0-h+12}}{Y_{T_0,M_0-h+12}} \tag{A20}$$

for $h = 0, 1, 2, ...$.

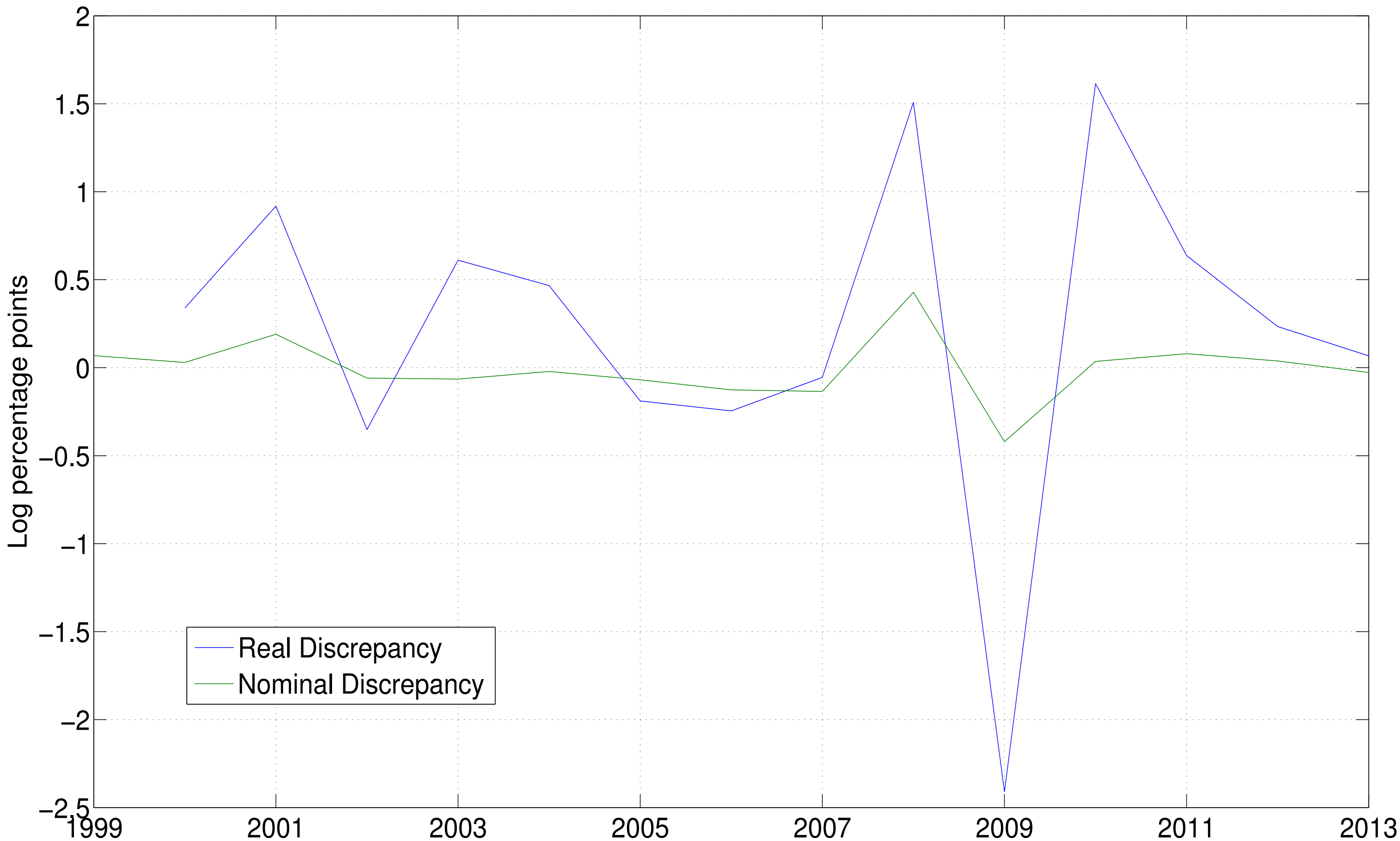

FIGURE 1. Discrepancy between average of 4-quarter log changes and annual (y/y) log changes.

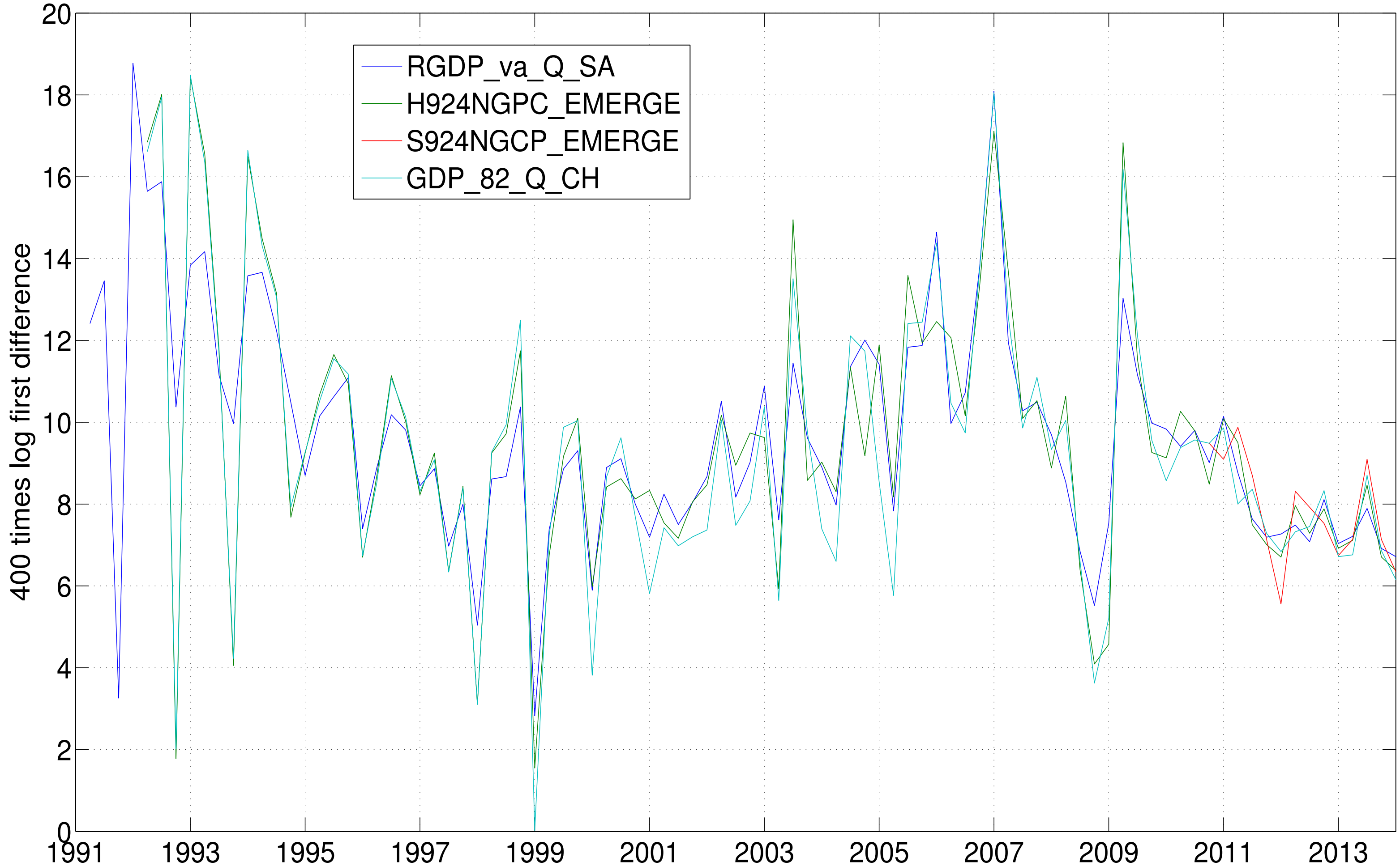

FIGURE 2. Alternative Measures of Real GDP-va. RGDP_va_Q_SA is our construction. H924NGPC_EMERGE_Q is from Haver Analytics. S924NGCP_EMERGE_Q is from NBS. GDP_82_Q_CH_FAMEQ is from Federal Reserve Board FAME database.

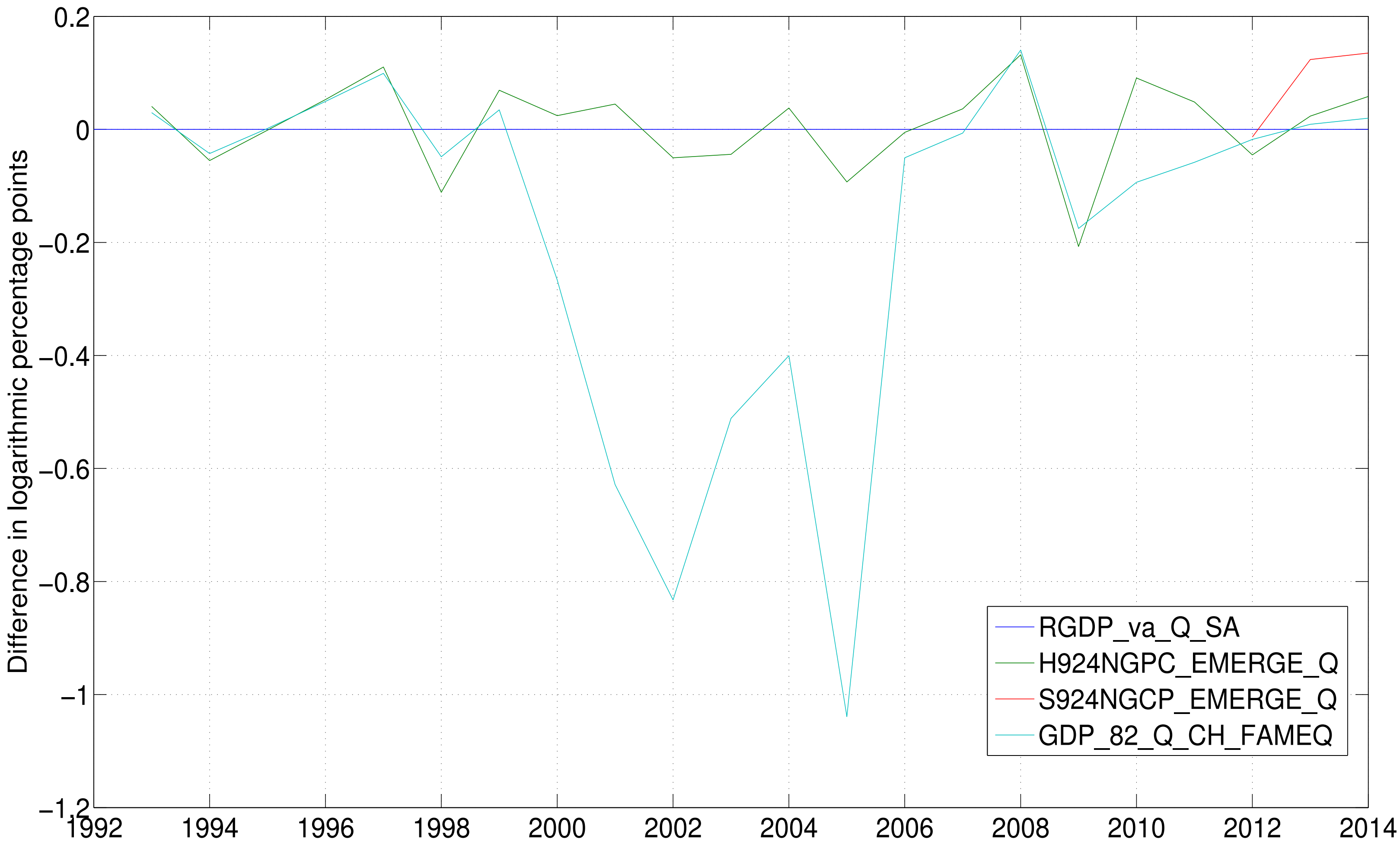

FIGURE 3. Difference in annual real GDP-va full year growth rates between aggregated quarterly measures and published annual measure. RGDP_va_Q_SA is our construction. H924NGPC_EMERGE_Q is from Haver Analytics. S924NGCP_EMERGE_Q is from NBS. GDP_82_Q_CH_FAMEQ is from Federal Reserve Board FAME database.

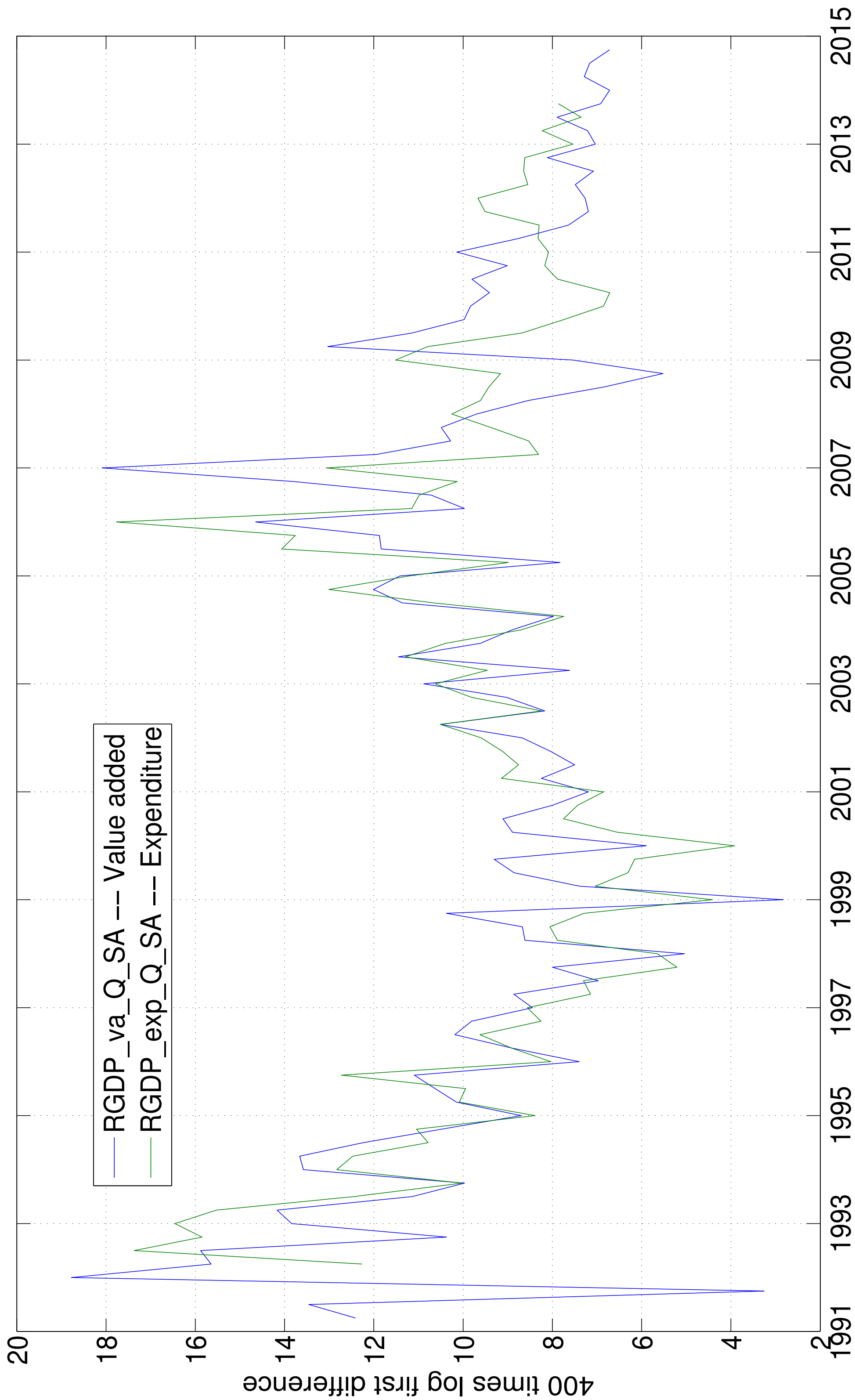

FIGURE 4. Our interpolations of real GDP-va and real GDP-exp.

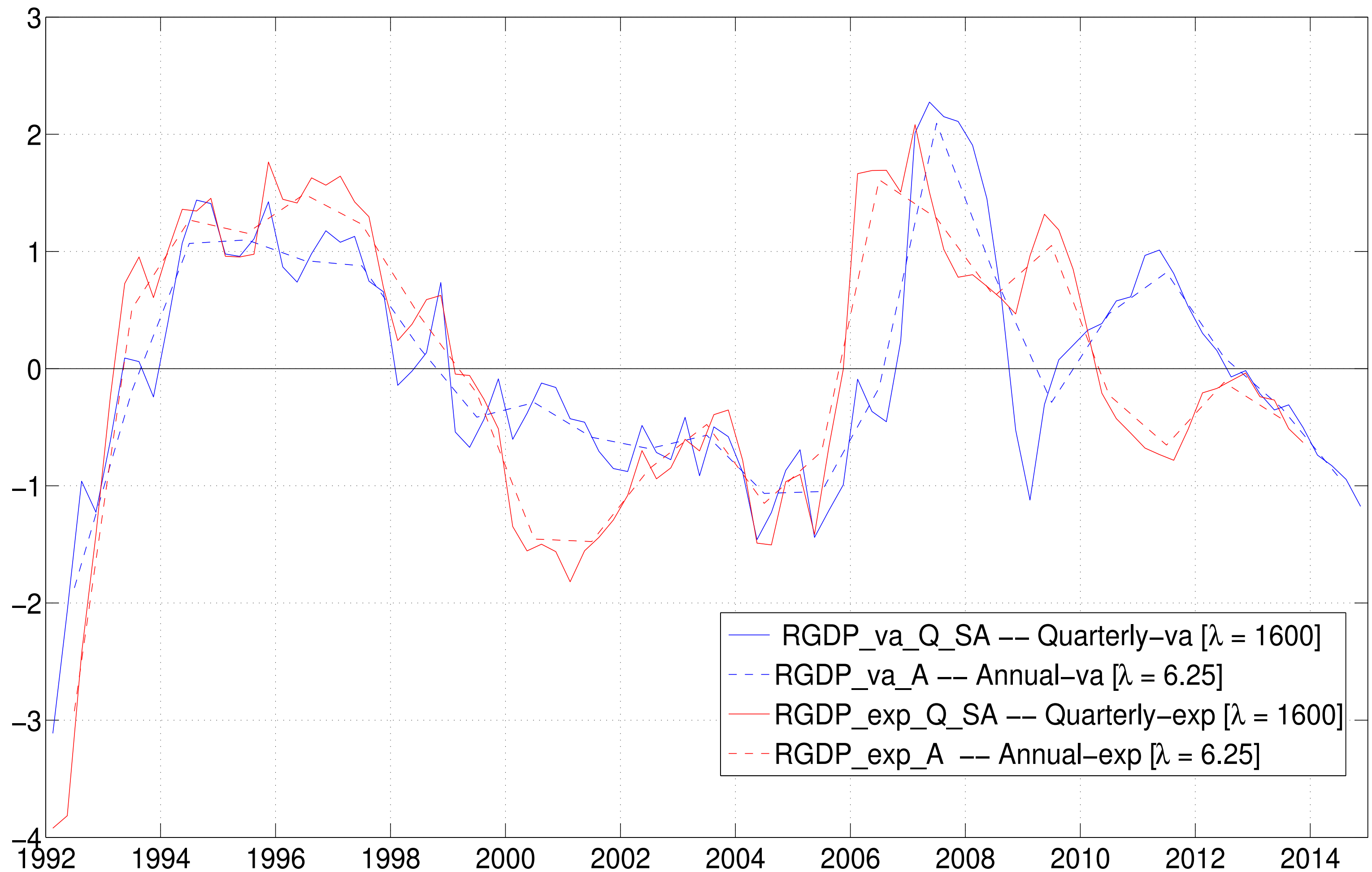

FIGURE 5. HP filtered cycles for annual and quarterly real GDP-va and real GDP-exp.

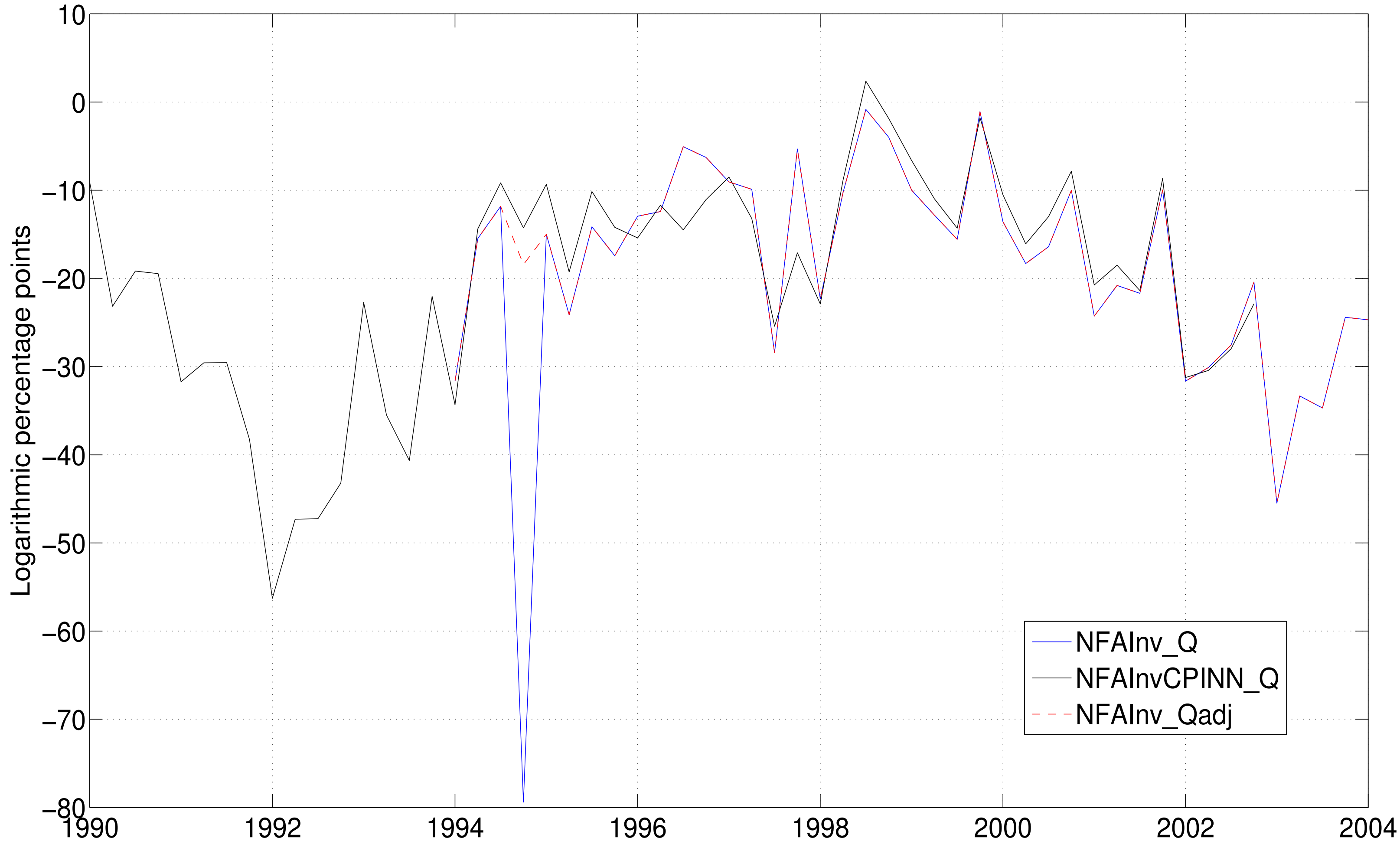

FIGURE 6. Negative of the 4-quarter logarithmic percent change lagged 4-quarters. NFAInv_Q is unadjusted total fixed asset investment. NFAInvCPINN_Q is fixed asset investment, Capital construction + Innovation. NFAInv_Qadj is total fixed asset investment, 1994q4 outlier adjusted.

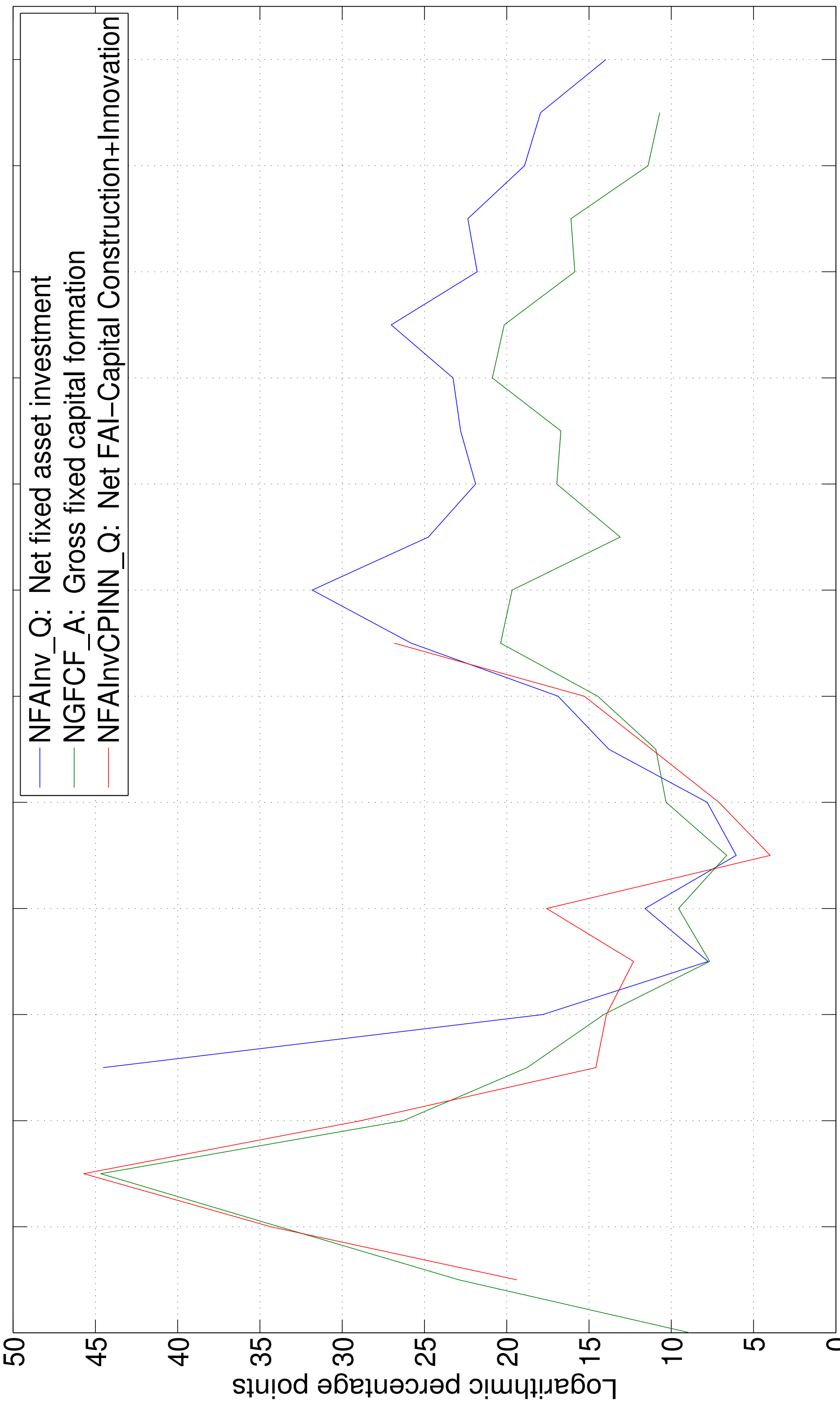

FIGURE 7. Full-year over full-year log percent change in investment measures.

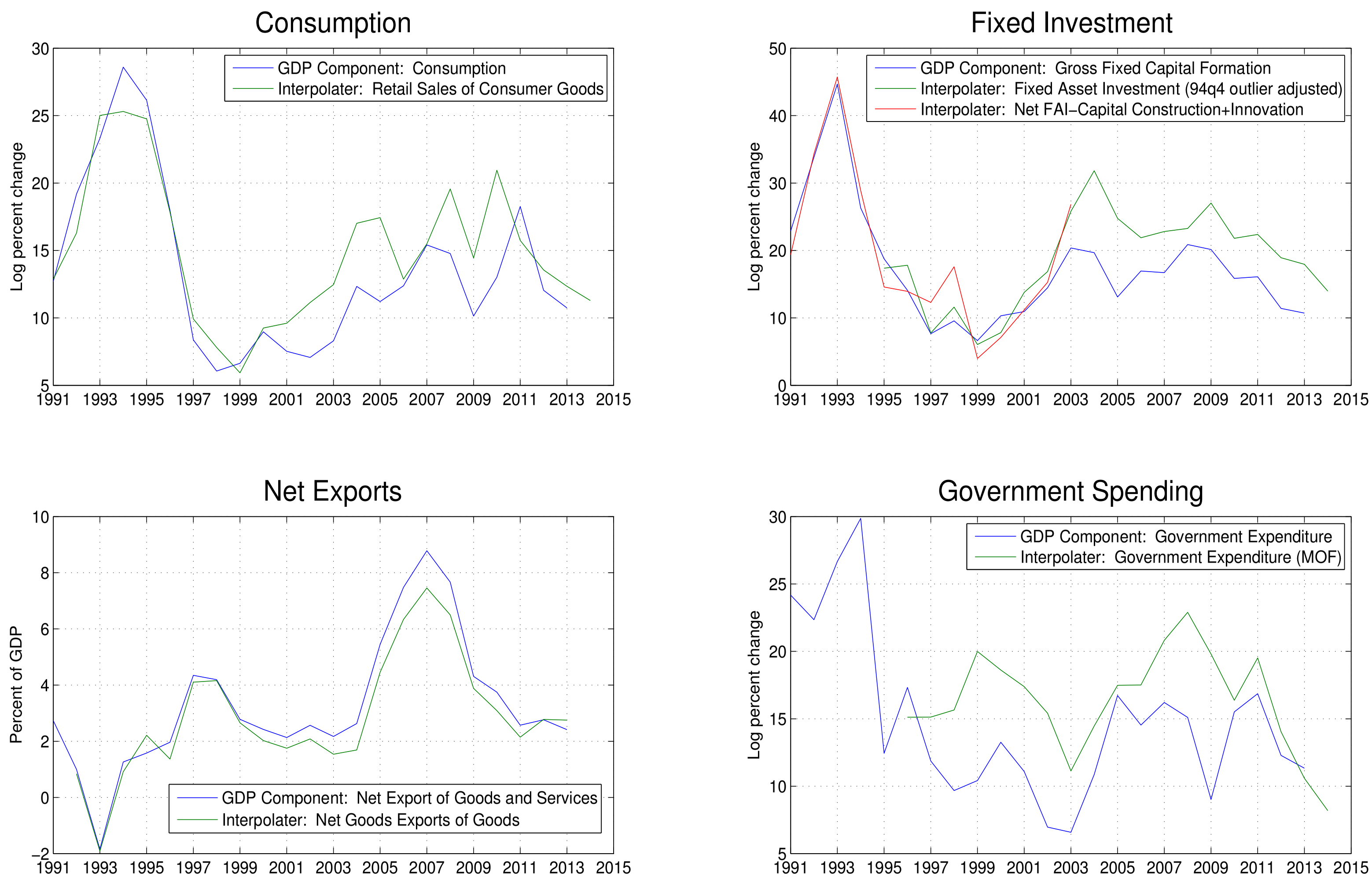

FIGURE 8. GDP subcomponents and interpolaters. Net exports is percent of GDP-exp. Remaining measures are full-year over full-year log percentage point changes.

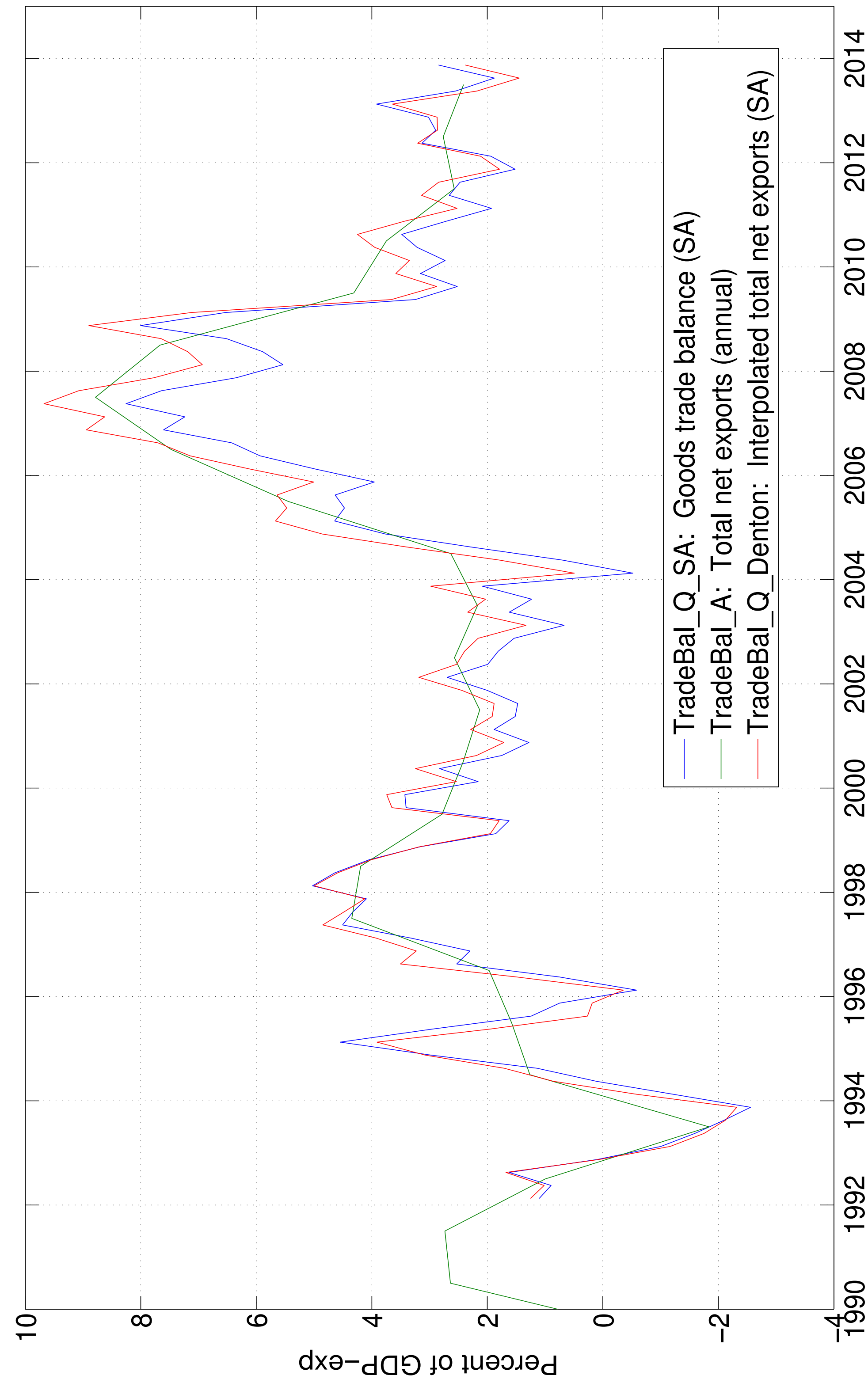

FIGURE 9. Measures of trade balance as percent of GDP-exp.

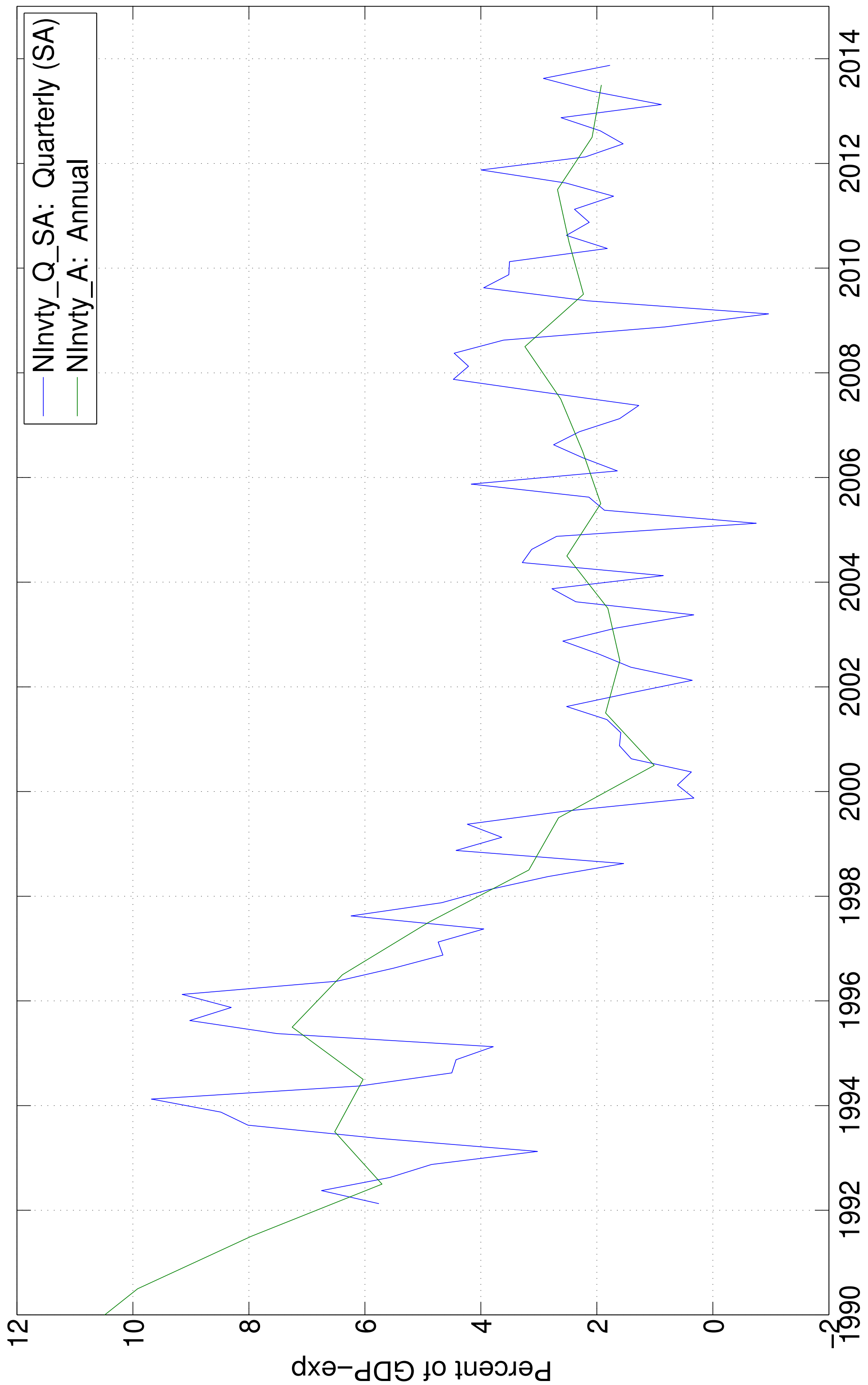

FIGURE 10. Change in private inventories as percentage of GDP-exp.

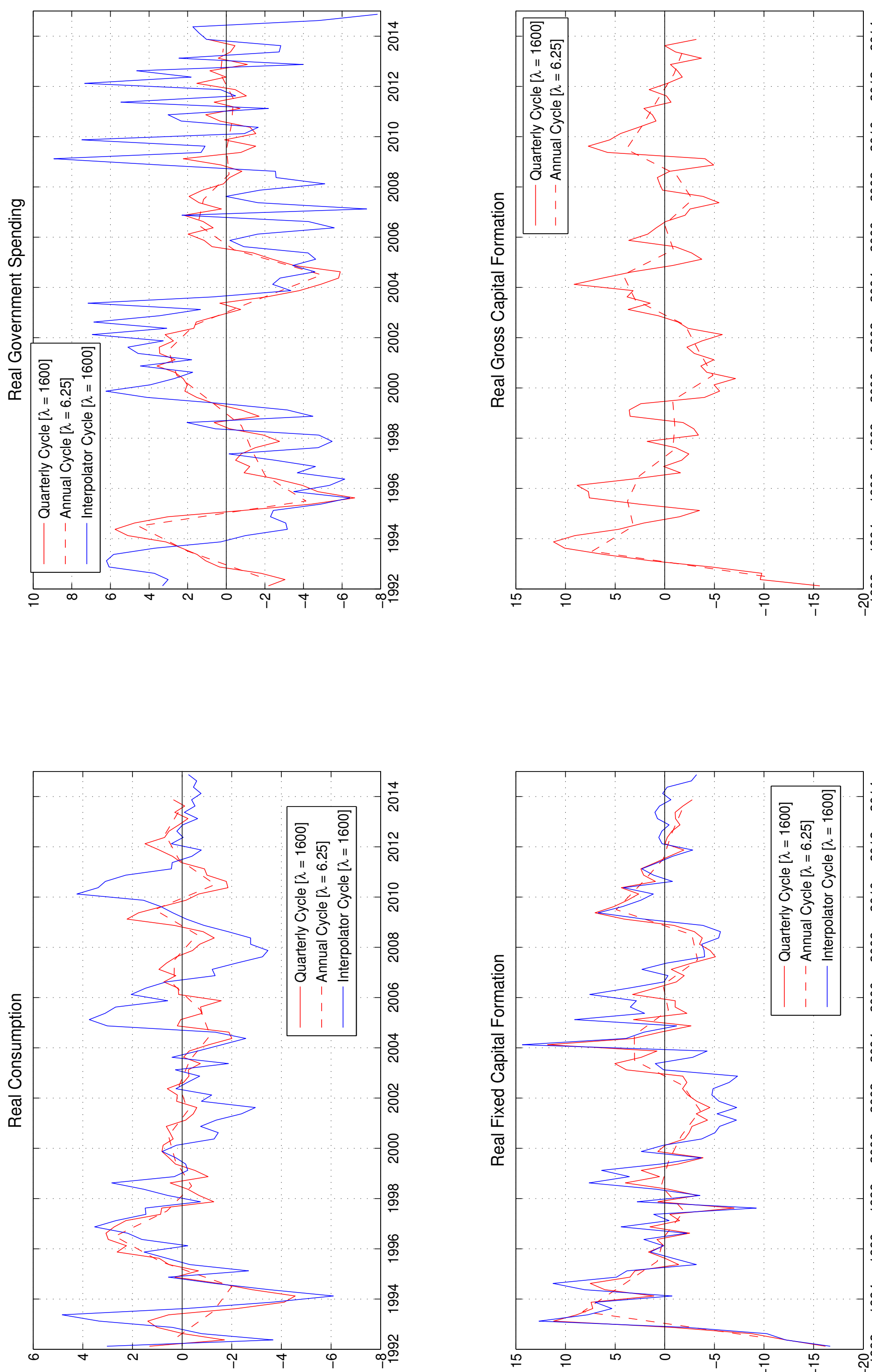

FIGURE 11. HP filter cycles of real GDP-exp subcomponents and interpolaters.

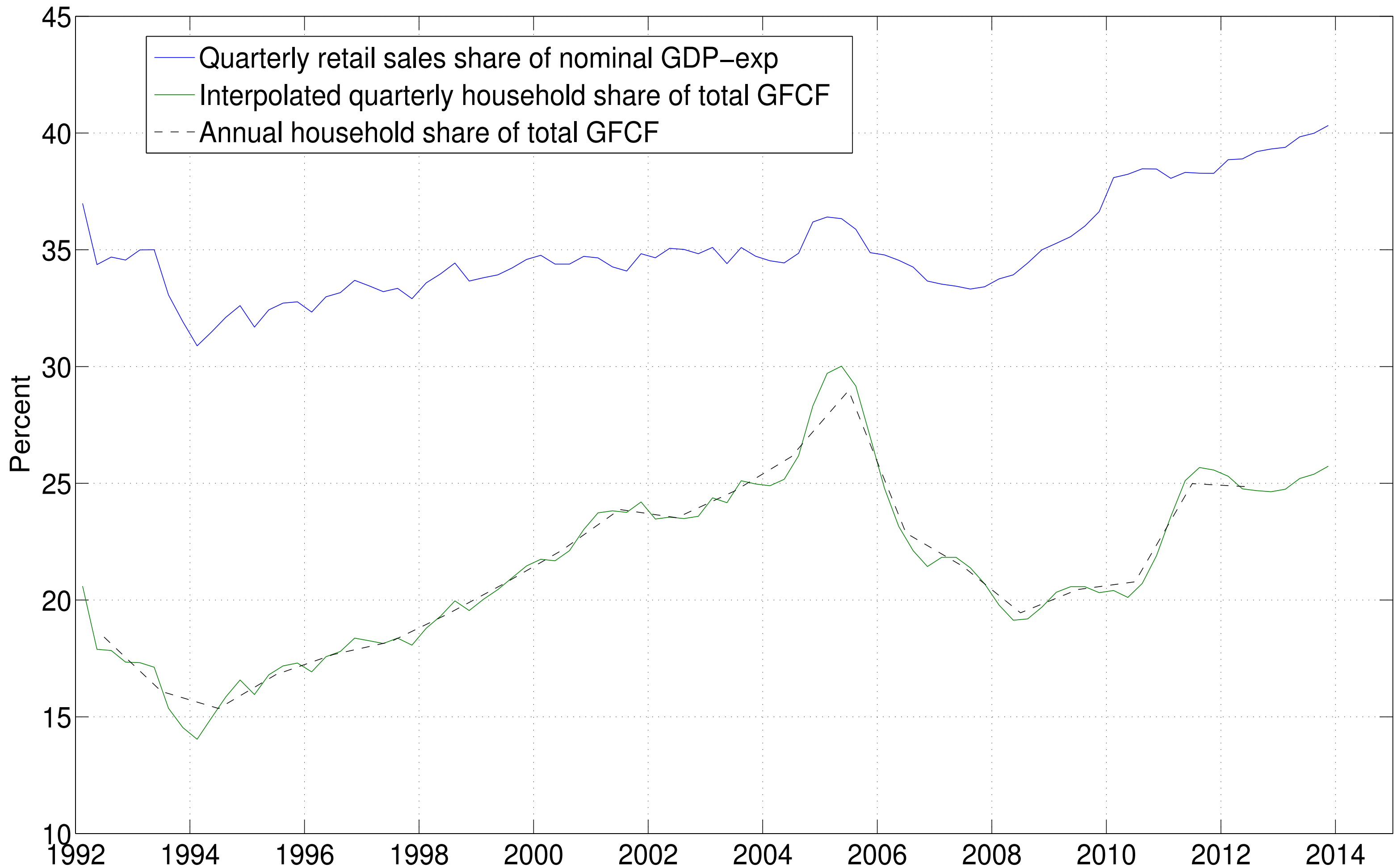

Figure 12. Interpolation of household share of total gross fixed capital formation (GFCF).

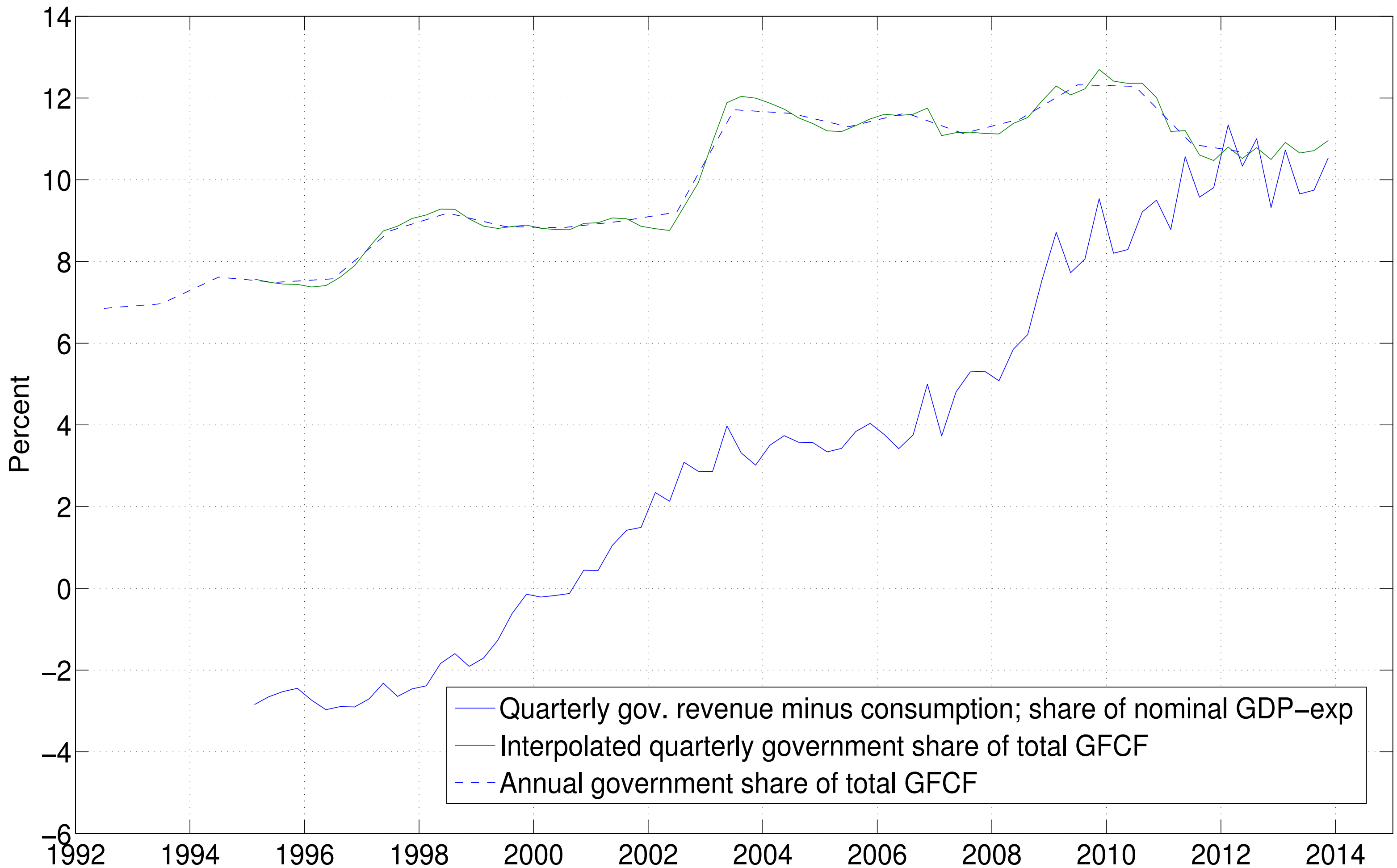

FIGURE 13. Interpolation of government share of total gross fixed capital formation (GFCF).

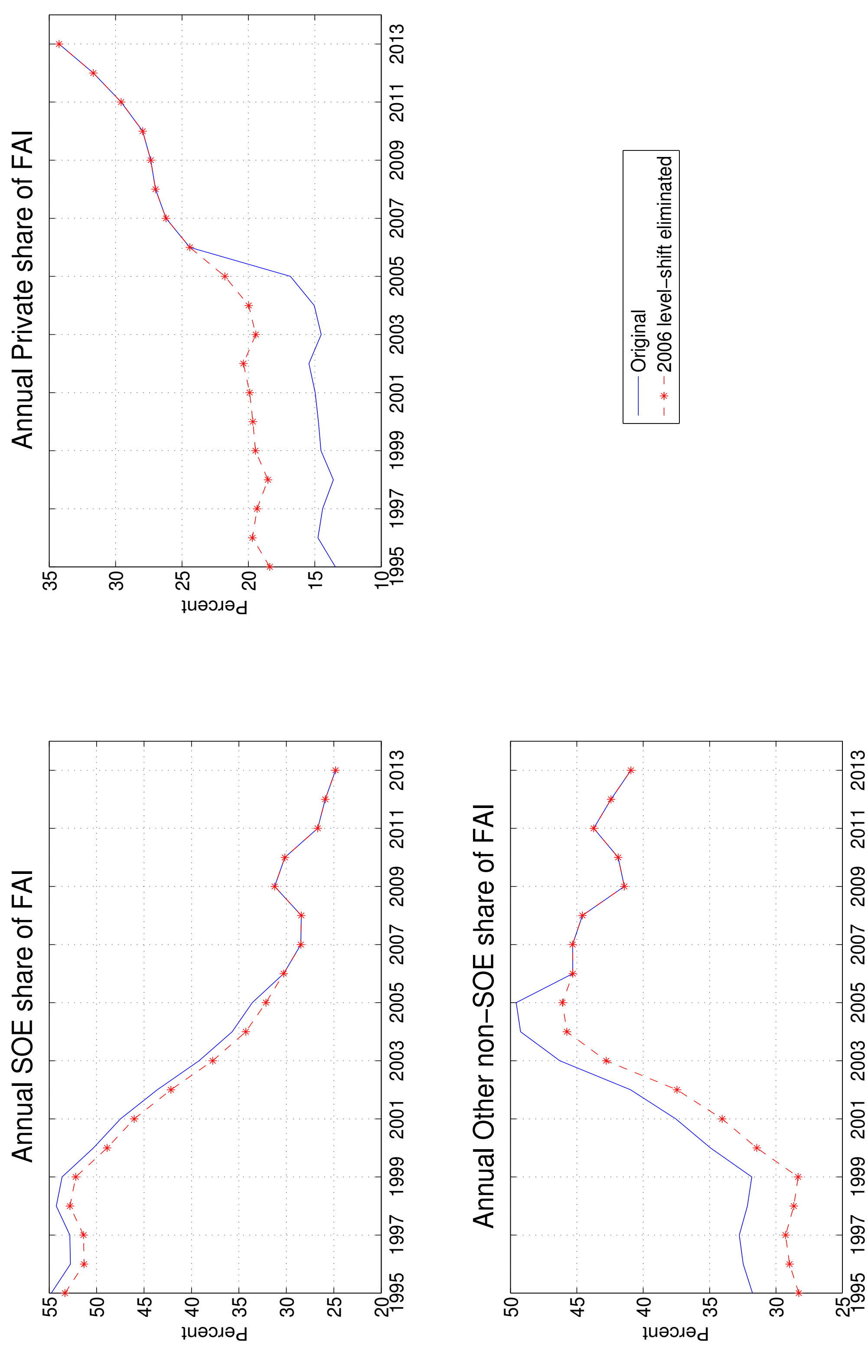

FIGURE 14. Annual sector shares of fixed assets investment (FAI).

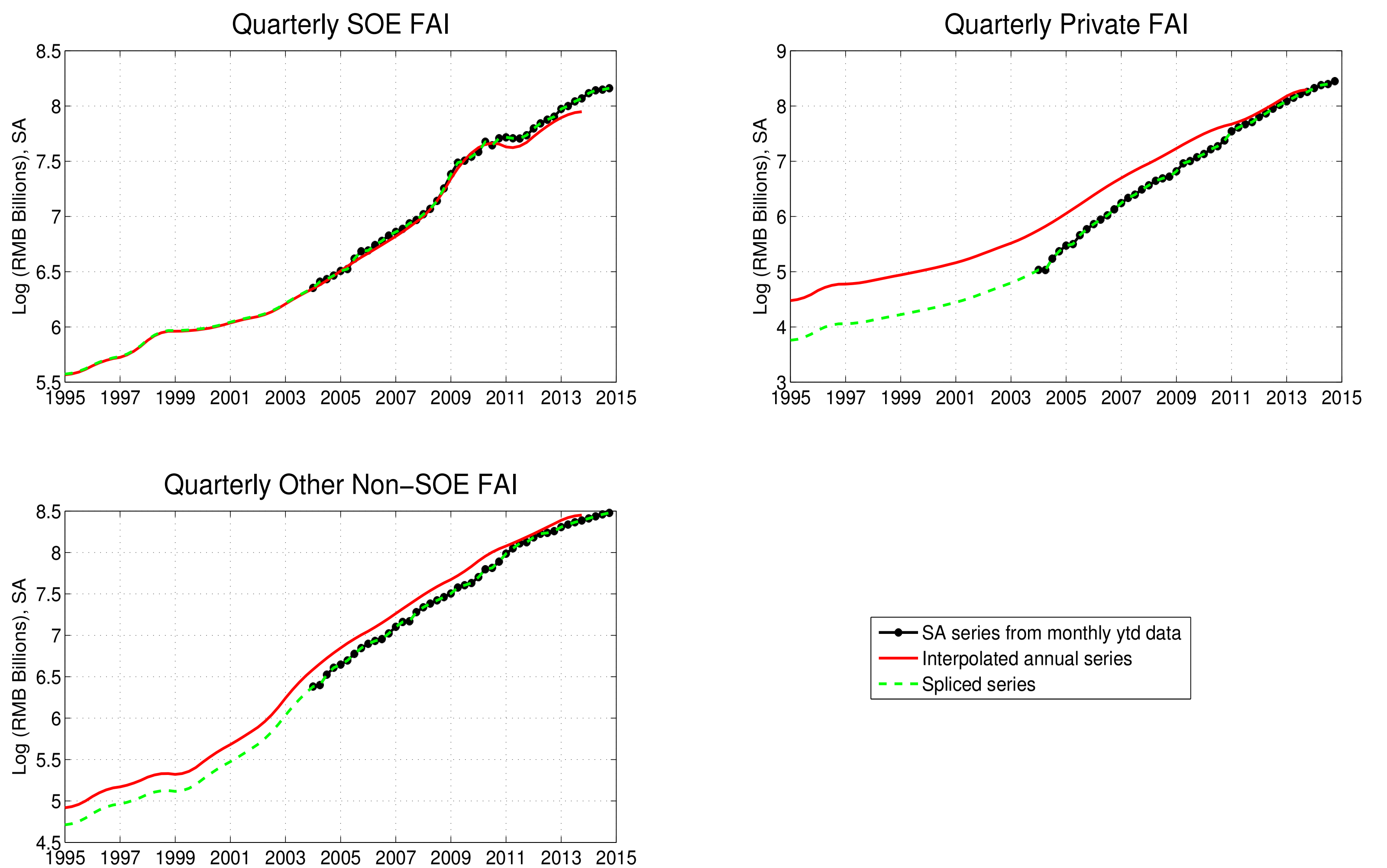

Figure 15. Quarterly measures of fixed assets investment (FAI) by sector. Black lines are SA series from CEIC monthly ytd data since 2004q1. Red lines are alternative annual series converted to quarterly frequency using minimal squared difference interpolation. Green series are black series spliced with red series.

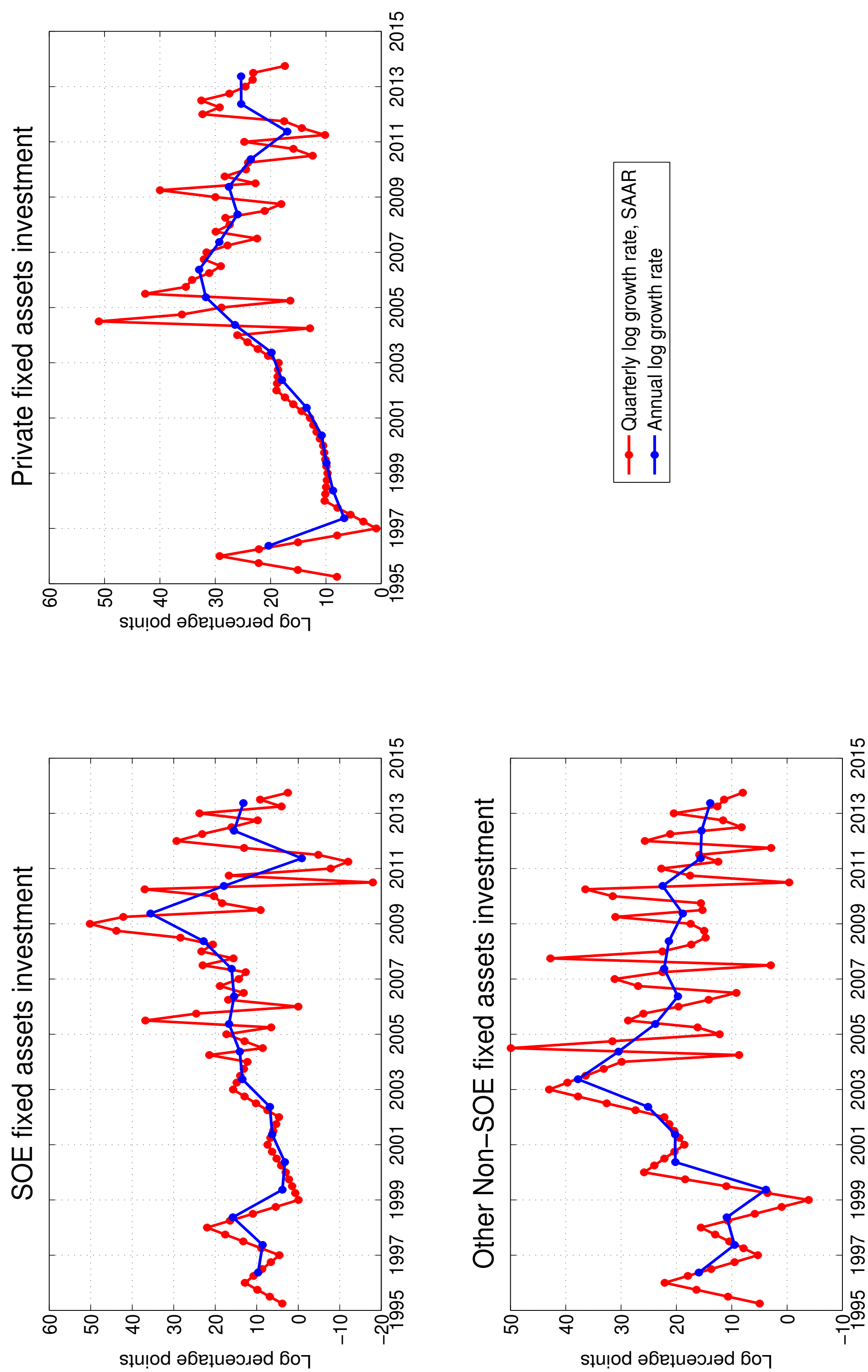

FIGURE 16. Growth rates of fixed assets investment (FAI) by sector.

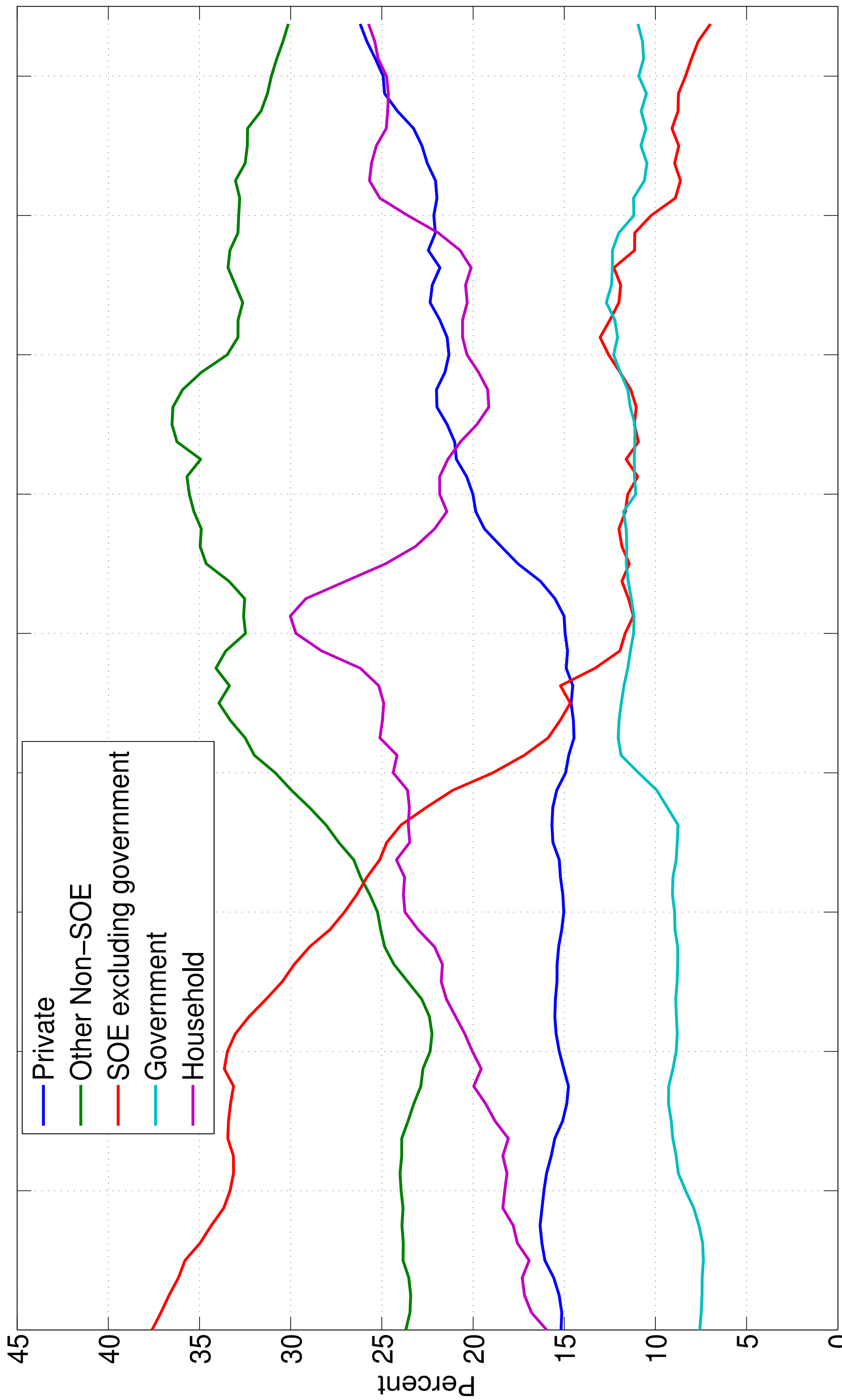

FIGURE 17. Quarterly sector shares of gross fixed capital formation.

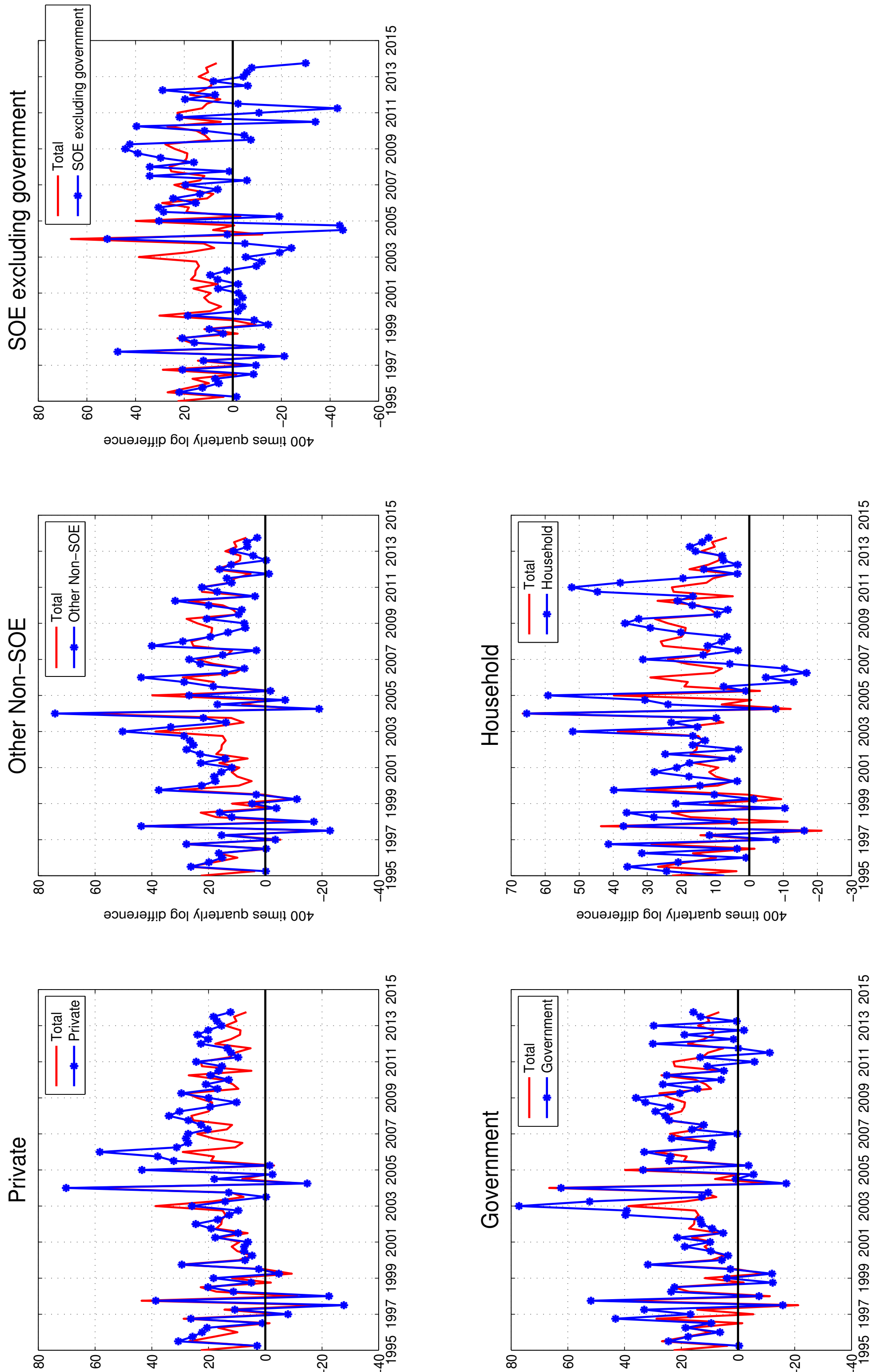

FIGURE 18. Quarterly growth rates of nominal gross fixed capital formation by sector.

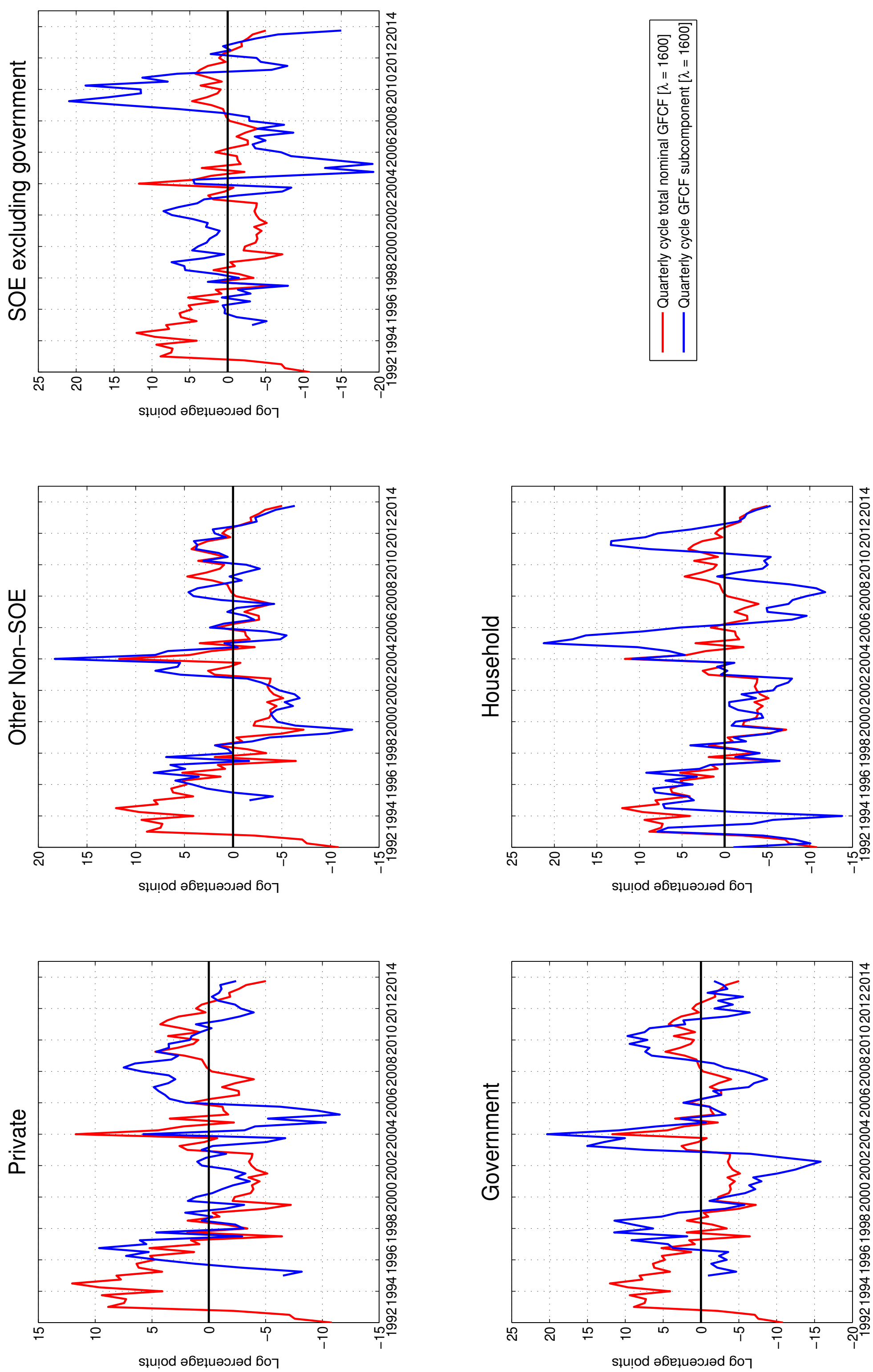

FIGURE 19. Quarterly HP cycles of nominal log gross fixed capital formation by sector.

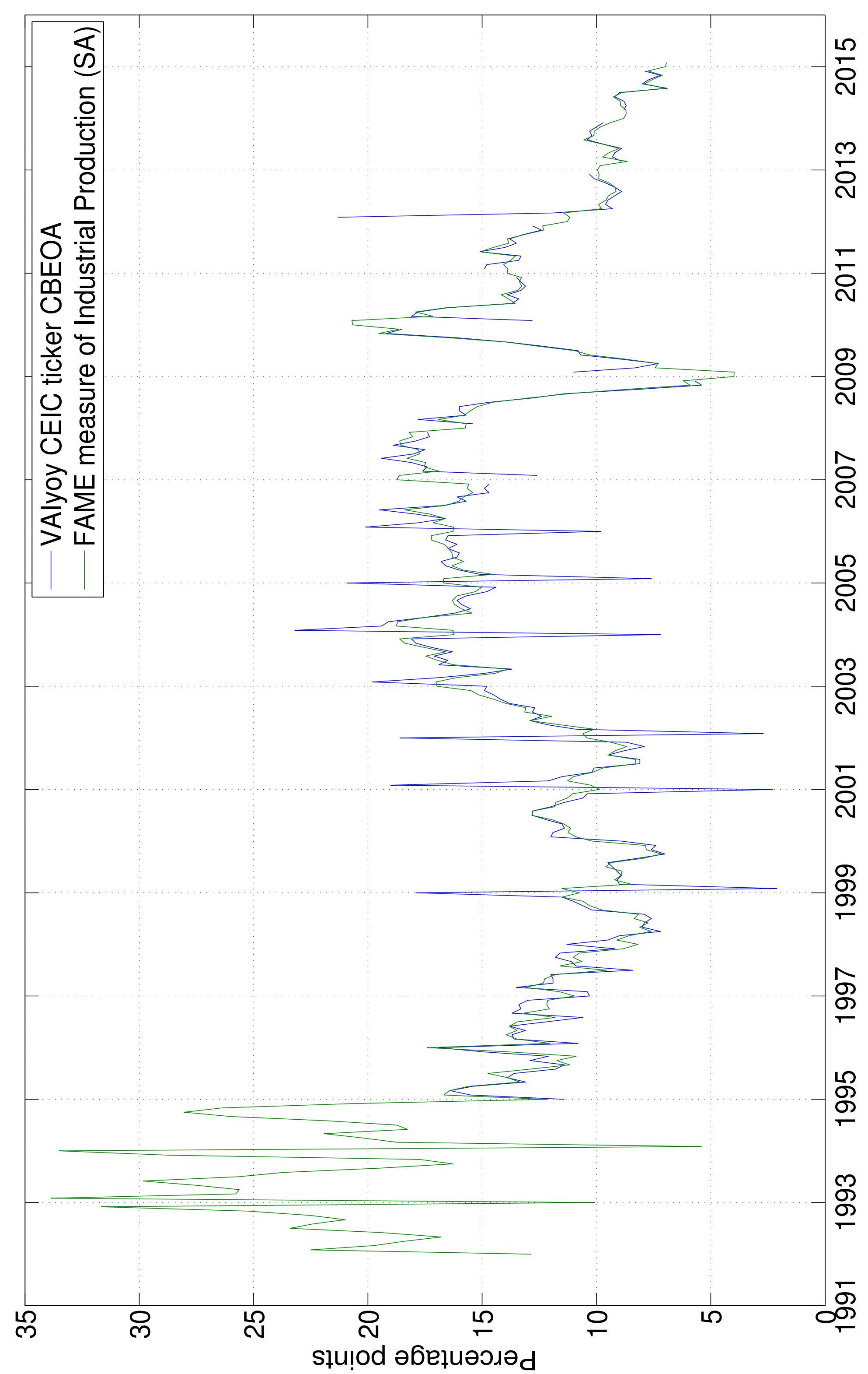

FIGURE 20. Alternative measures of nominal industrial production.

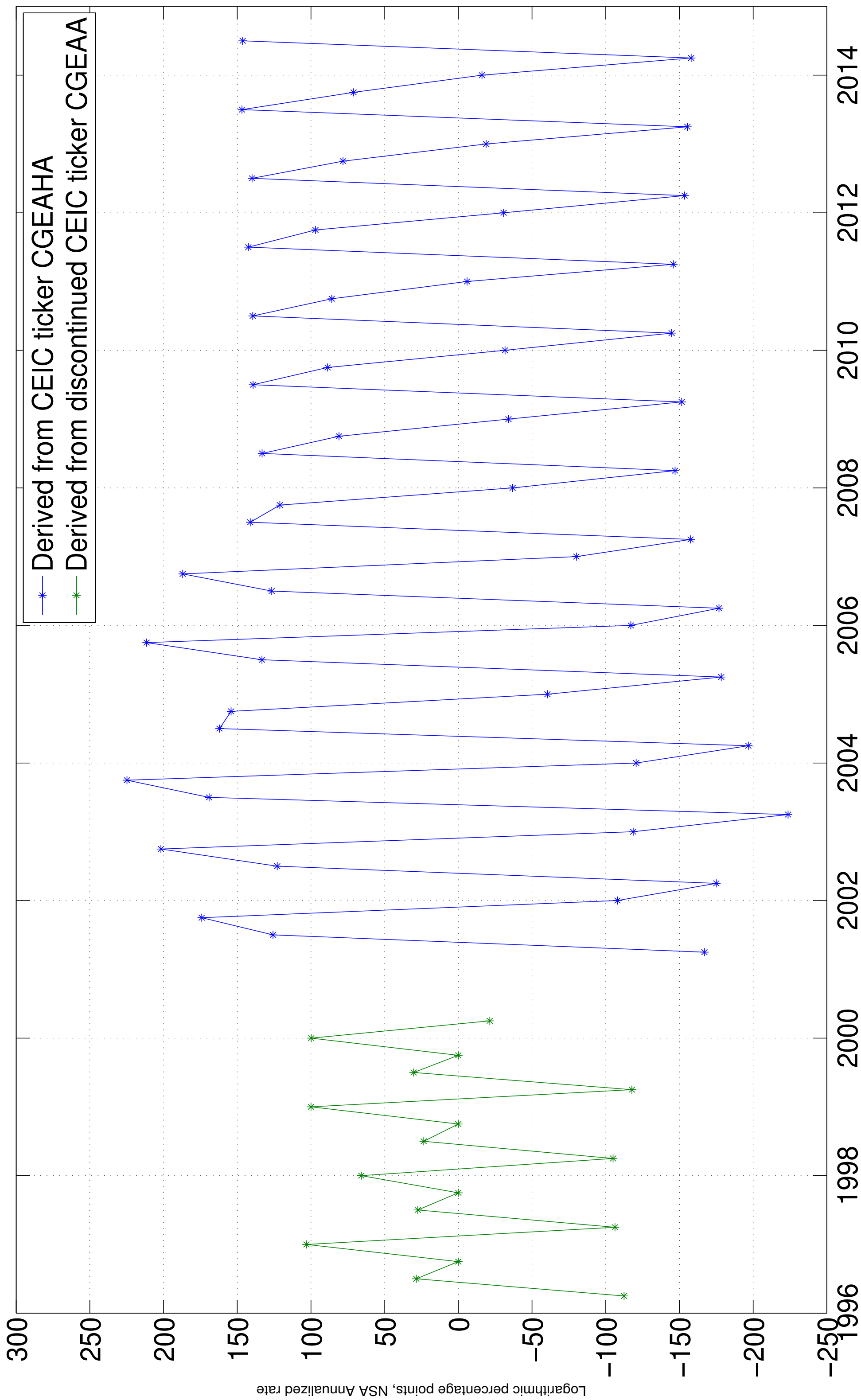

FIGURE 21. Quarterly growth rates of rural cash income per-capita (NSA).

FIGURE 22. Annual and quarterly rural shares of the urban population

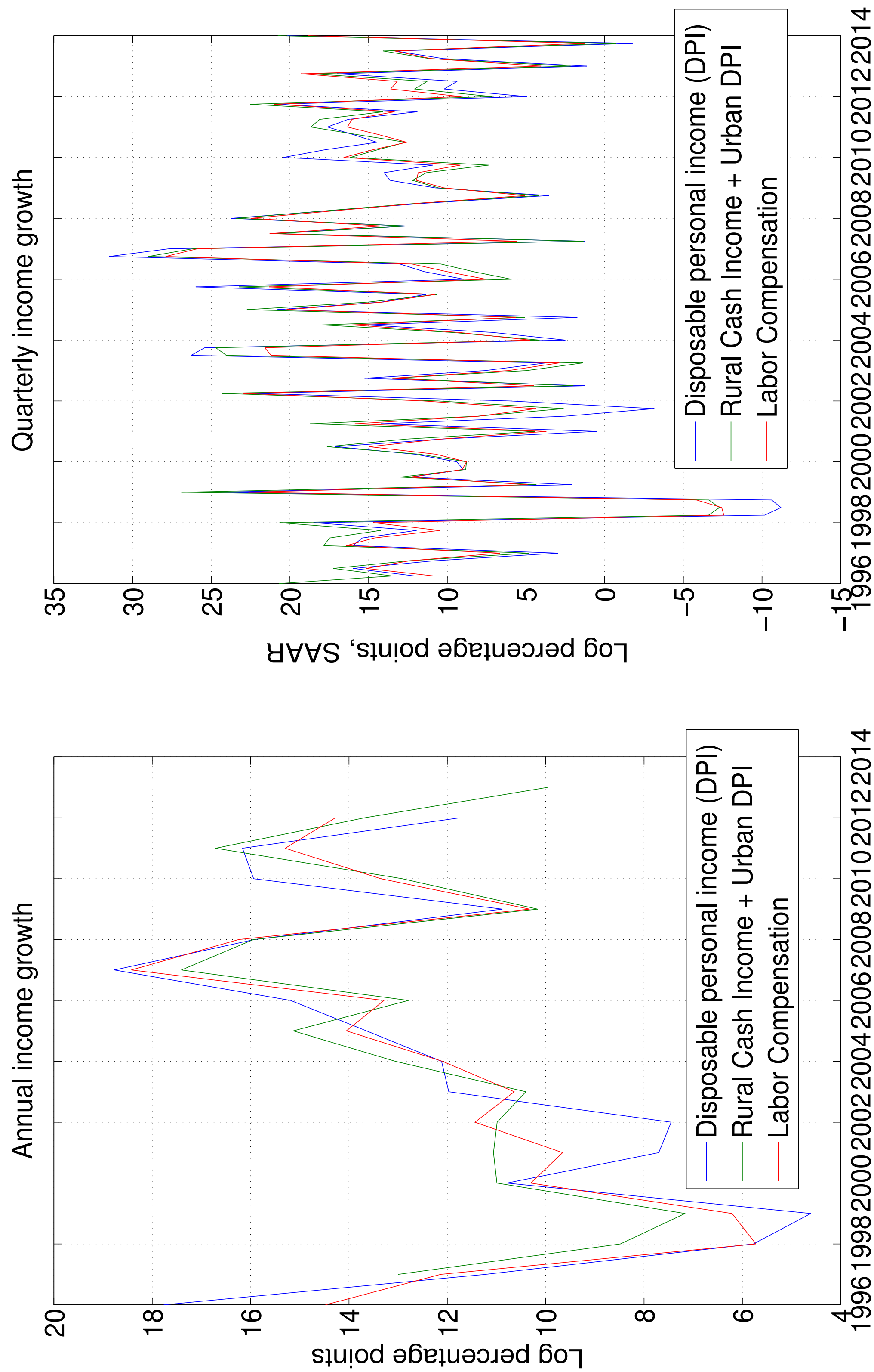

FIGURE 23. Alternative measures of aggregate labor income growth.

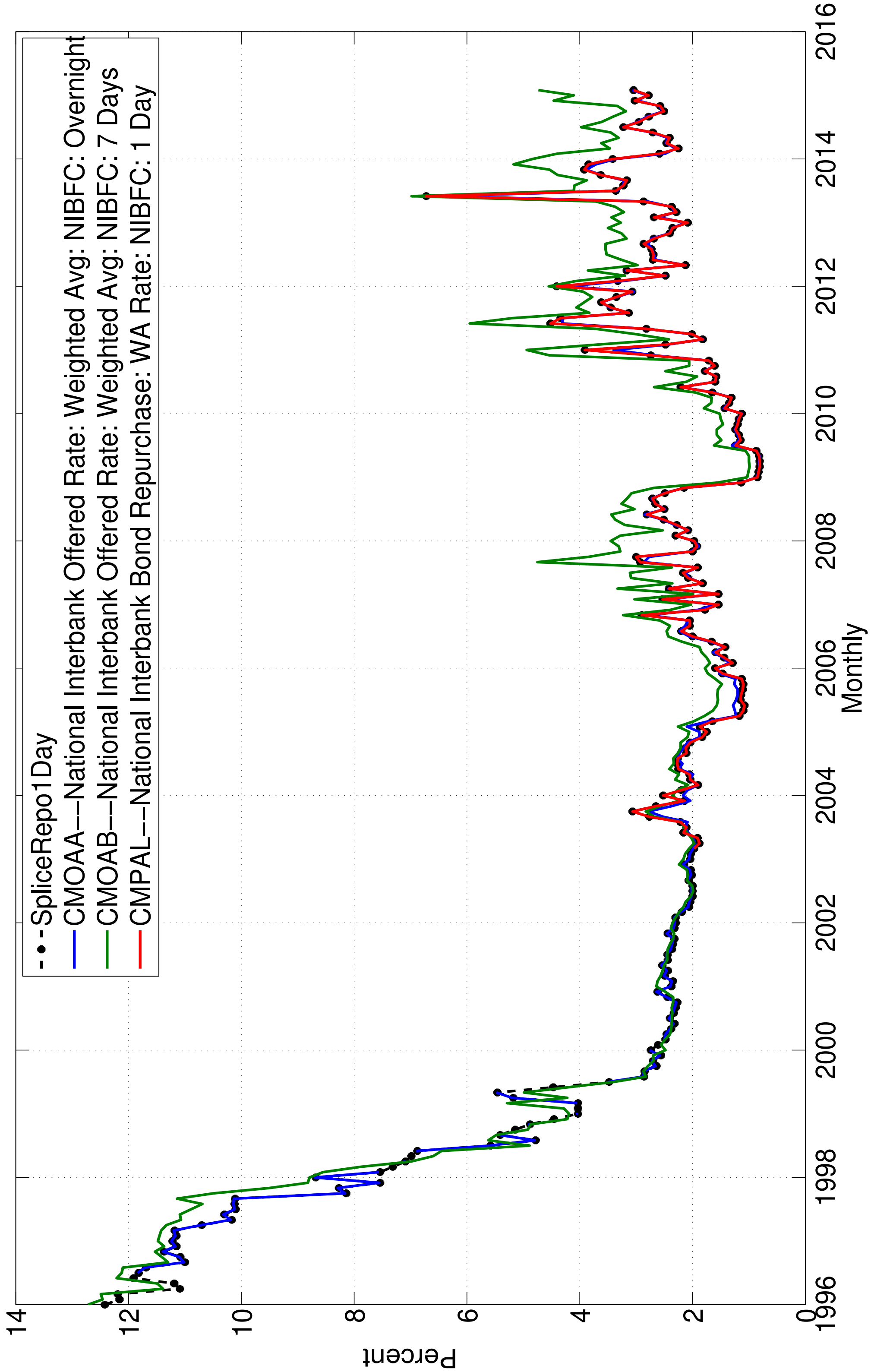

FIGURE 24. Components used to construct central bank policy rate [SpliceRepo1Day].

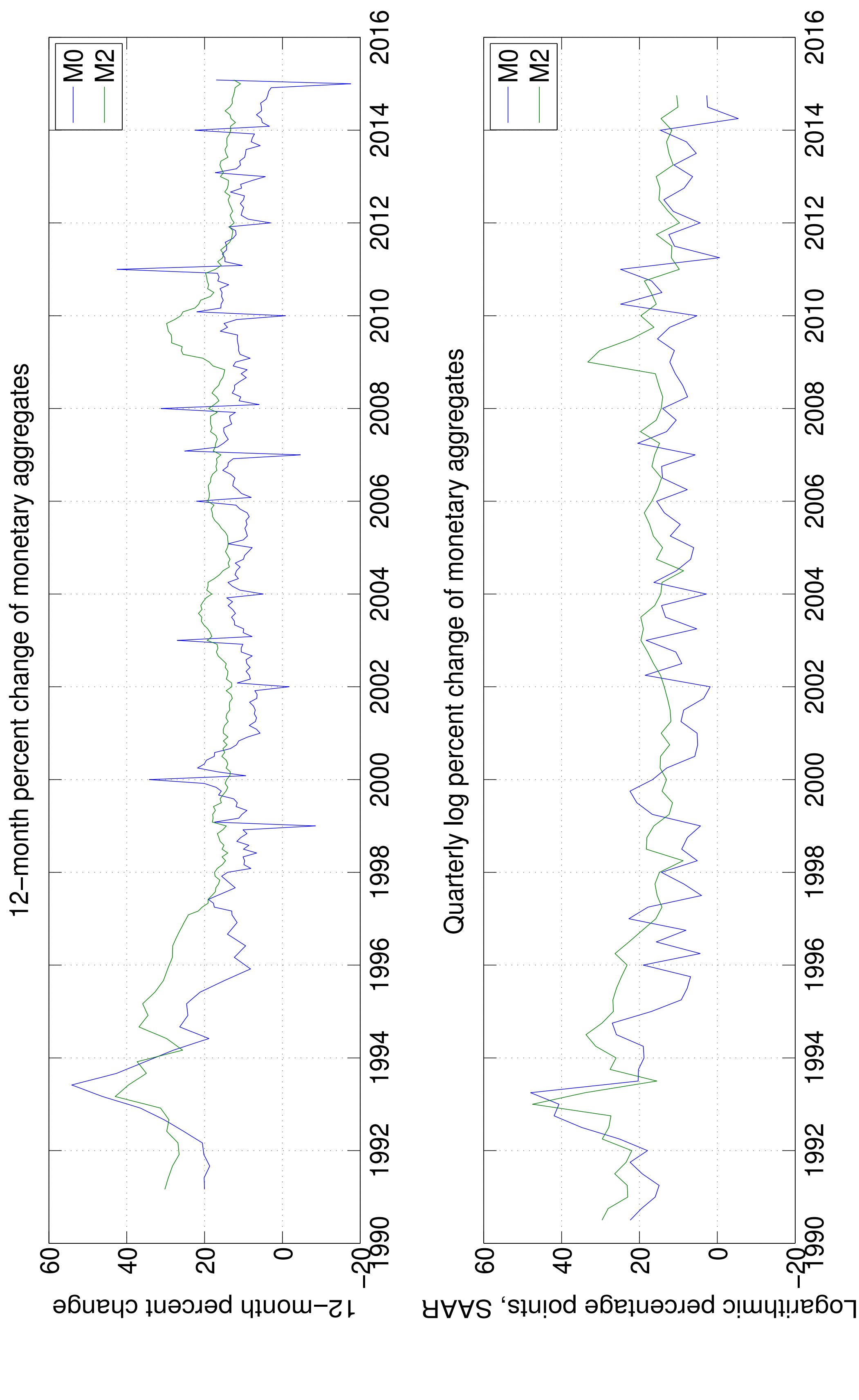

FIGURE 25. Growth rates of monetary aggregates.

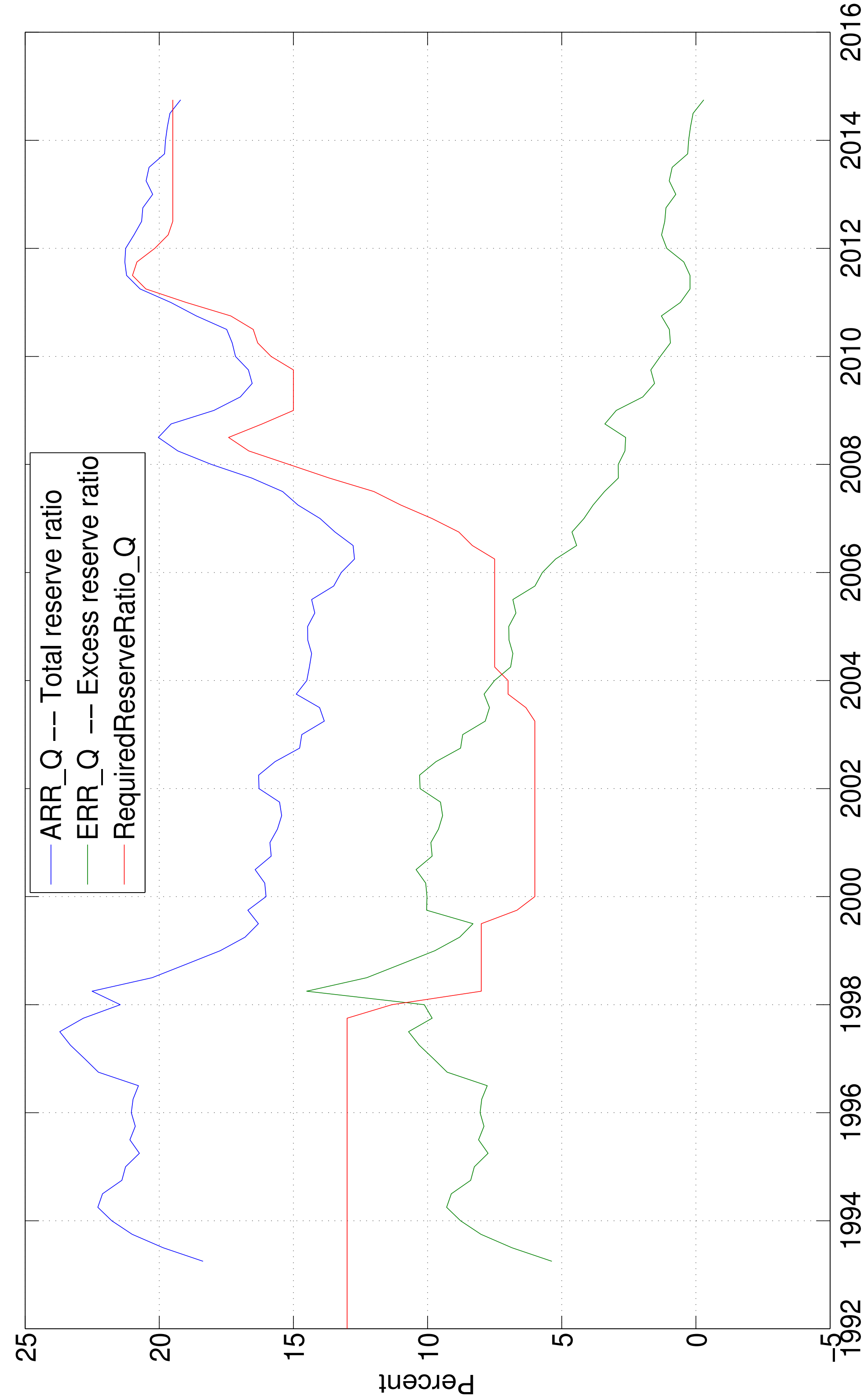

Figure 26. Alternative measures of reserve ratios, quarterly.

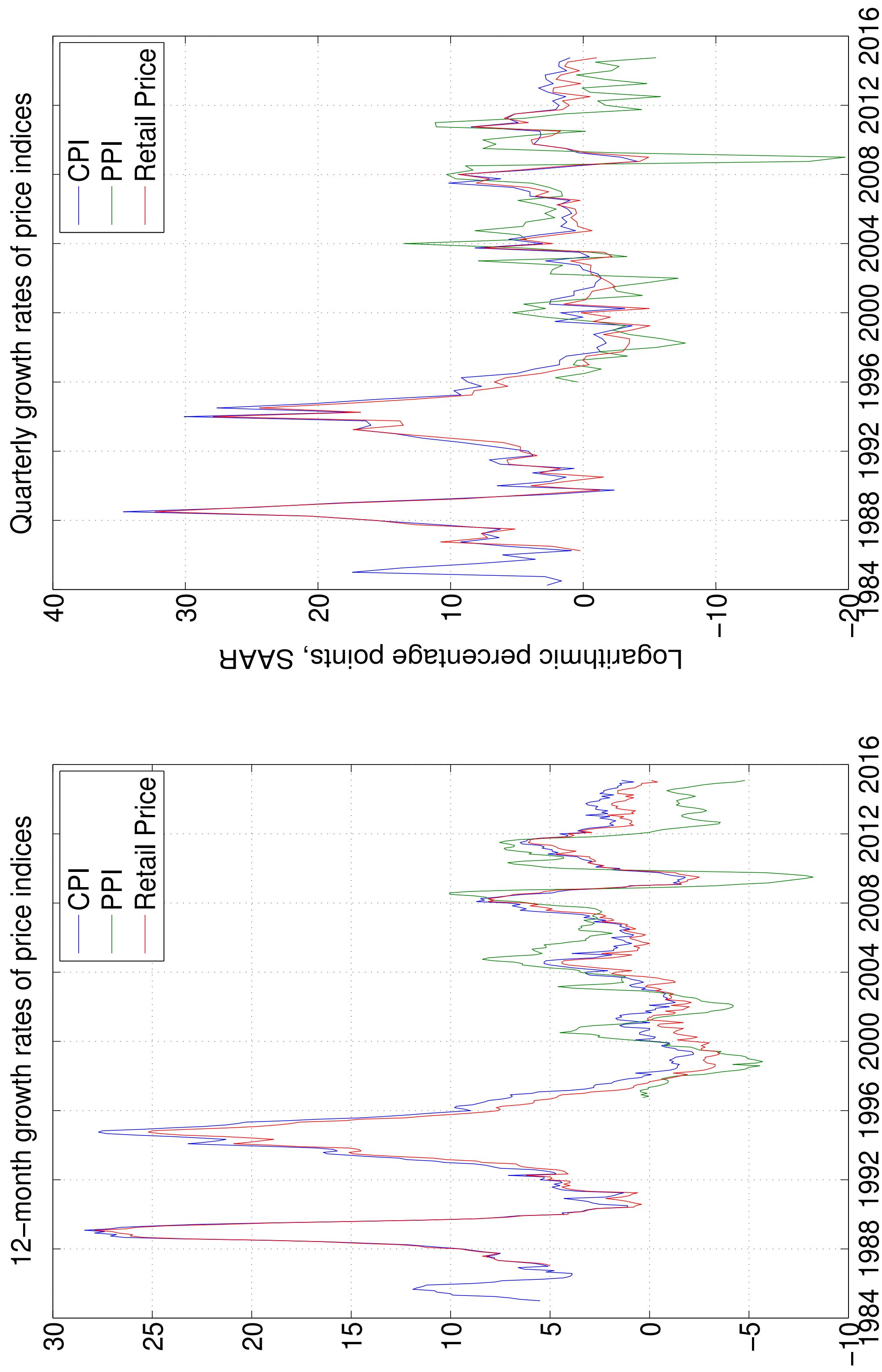

FIGURE 27. Alternative measures of inflation.

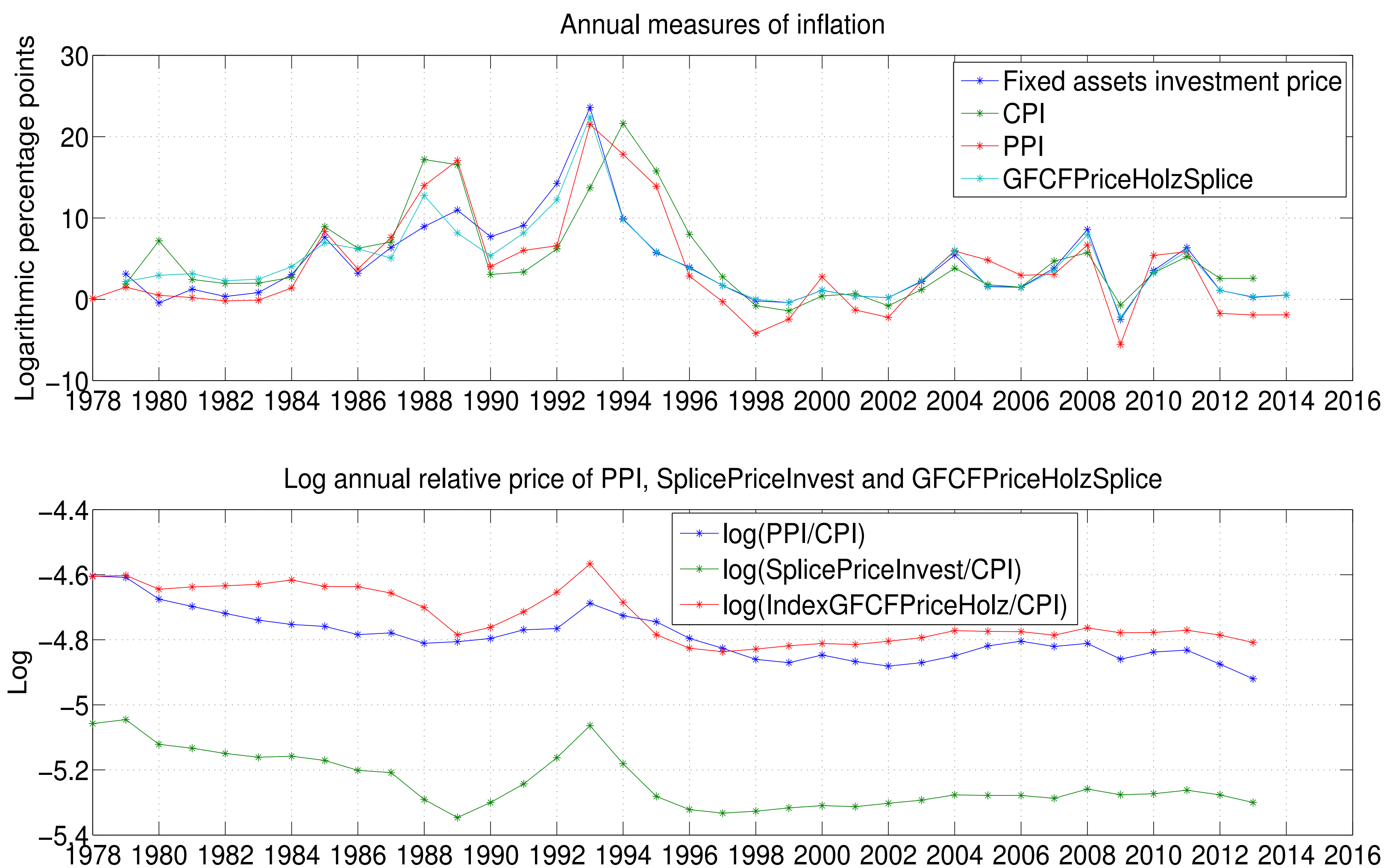

FIGURE 28. Annual inflation rates and relative prices for investment and other price indices.

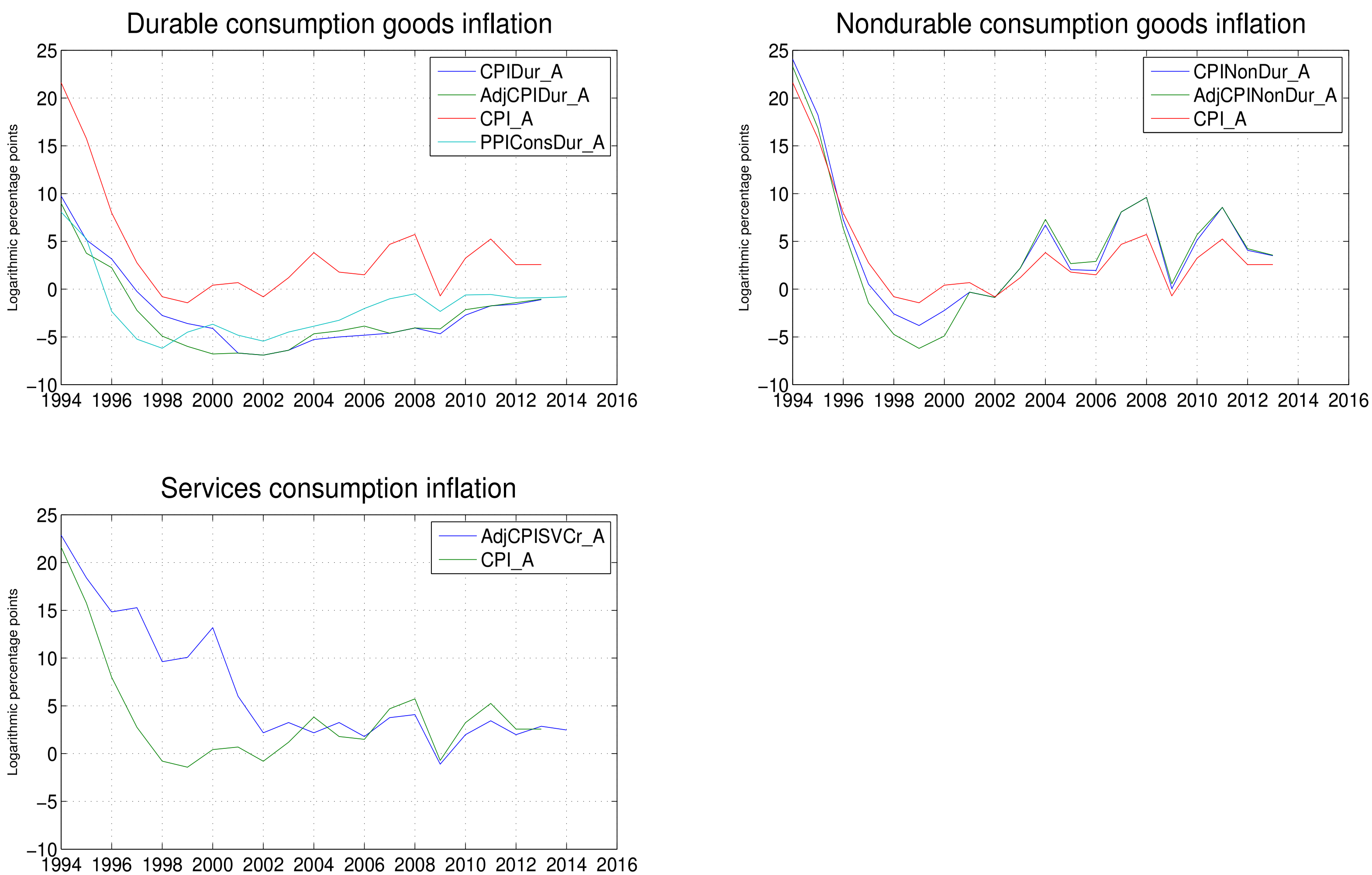

FIGURE 29. Annual inflation rates for durable, nondurable and services consumer prices.

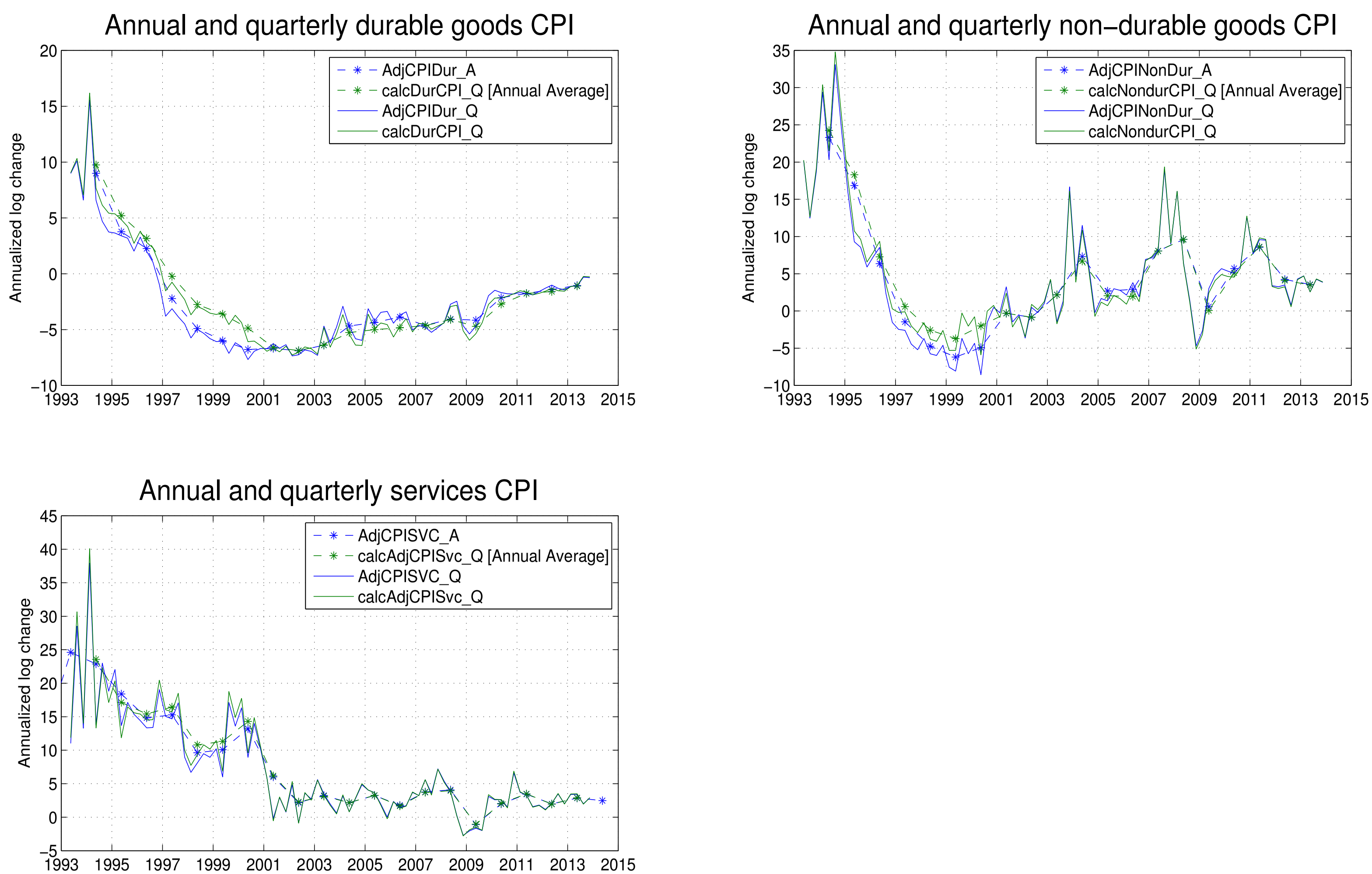

FIGURE 30. Annual and quarterly inflation rates for durable, nondurable and services consumer price indices.

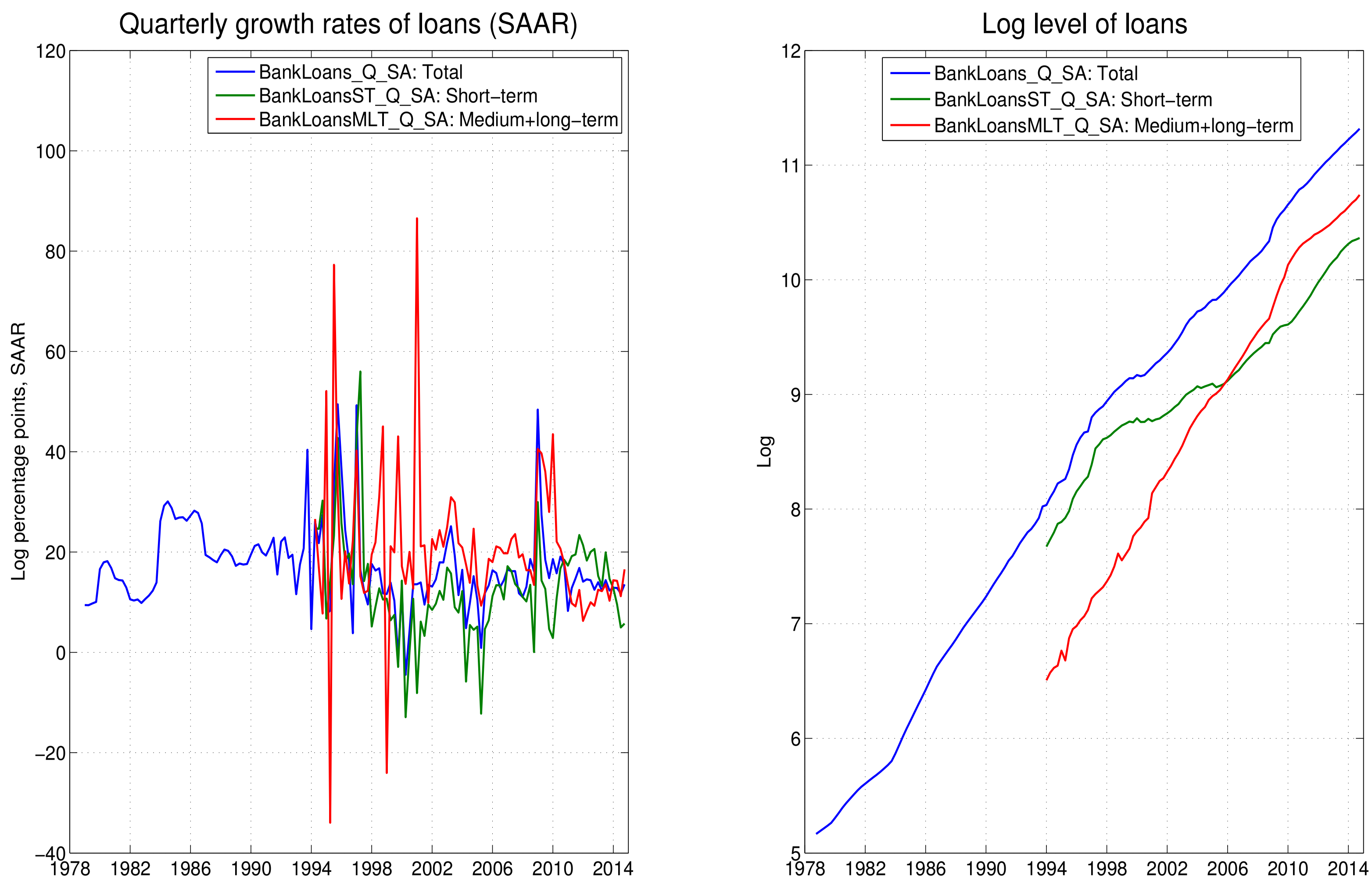

FIGURE 31. Measures of outstanding bank loans.

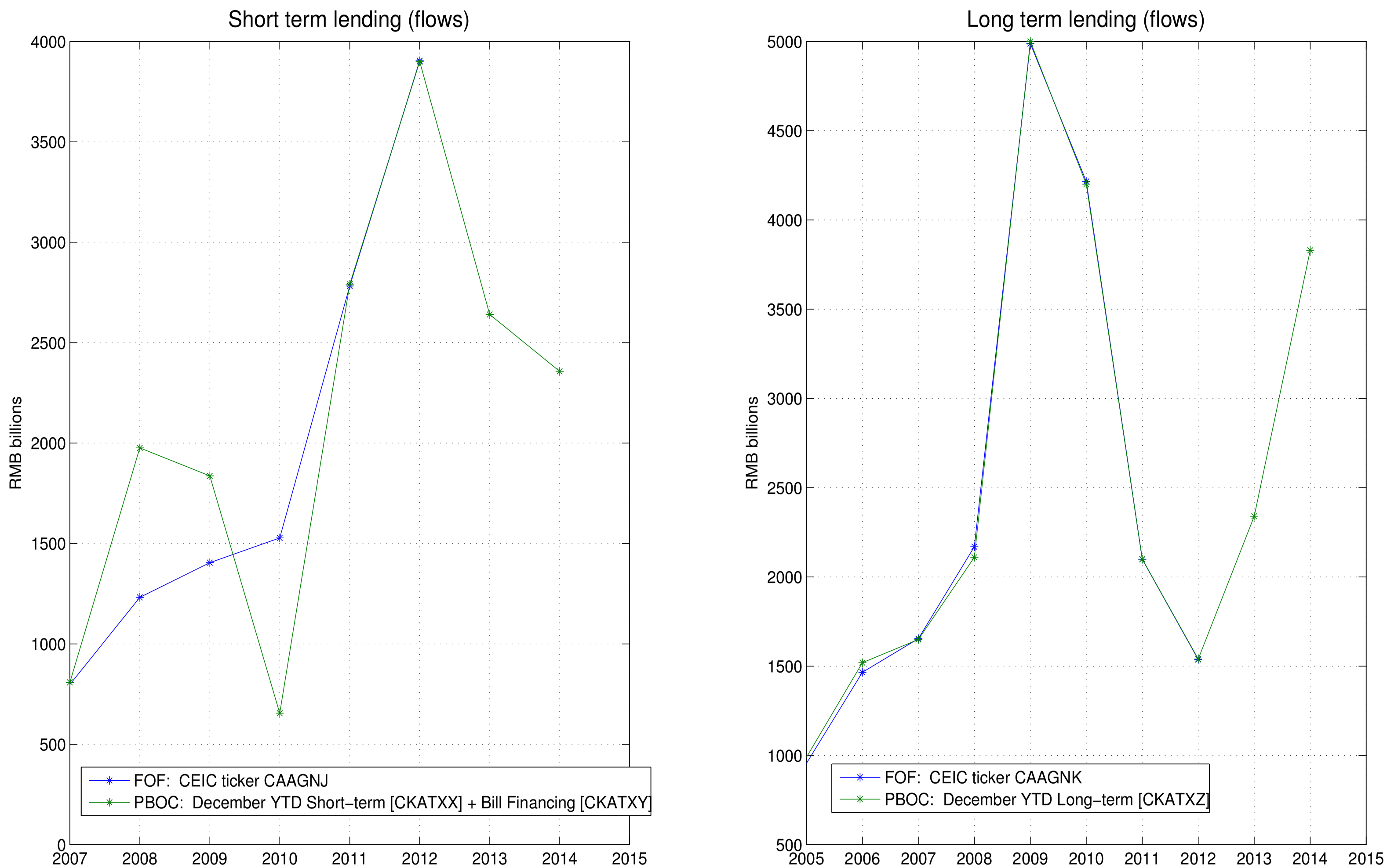

FIGURE 32. Alternative annual measures of new bank loans (flows).

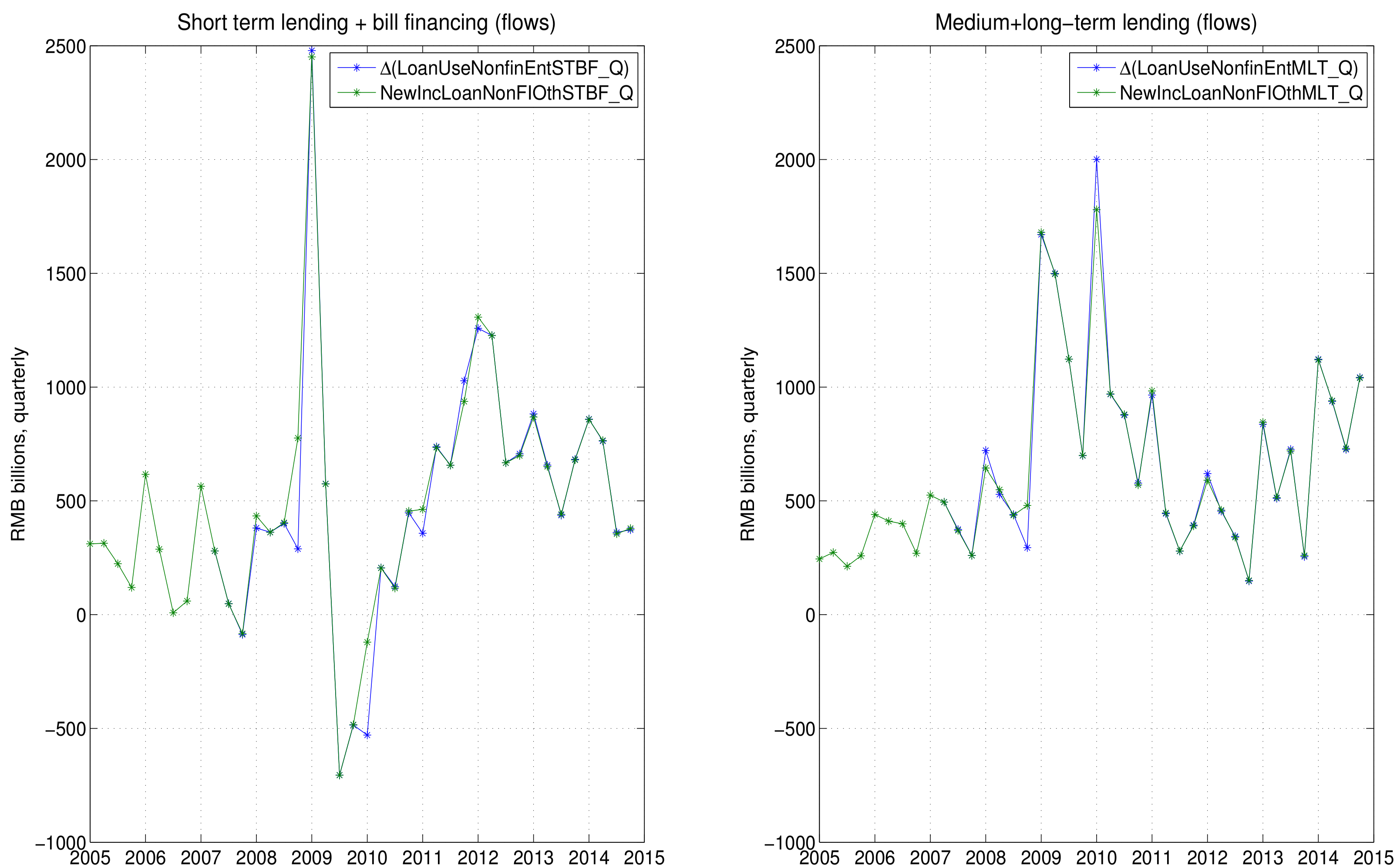

FIGURE 33. Alternative quarterly measures of flows of bank loans since 2005.

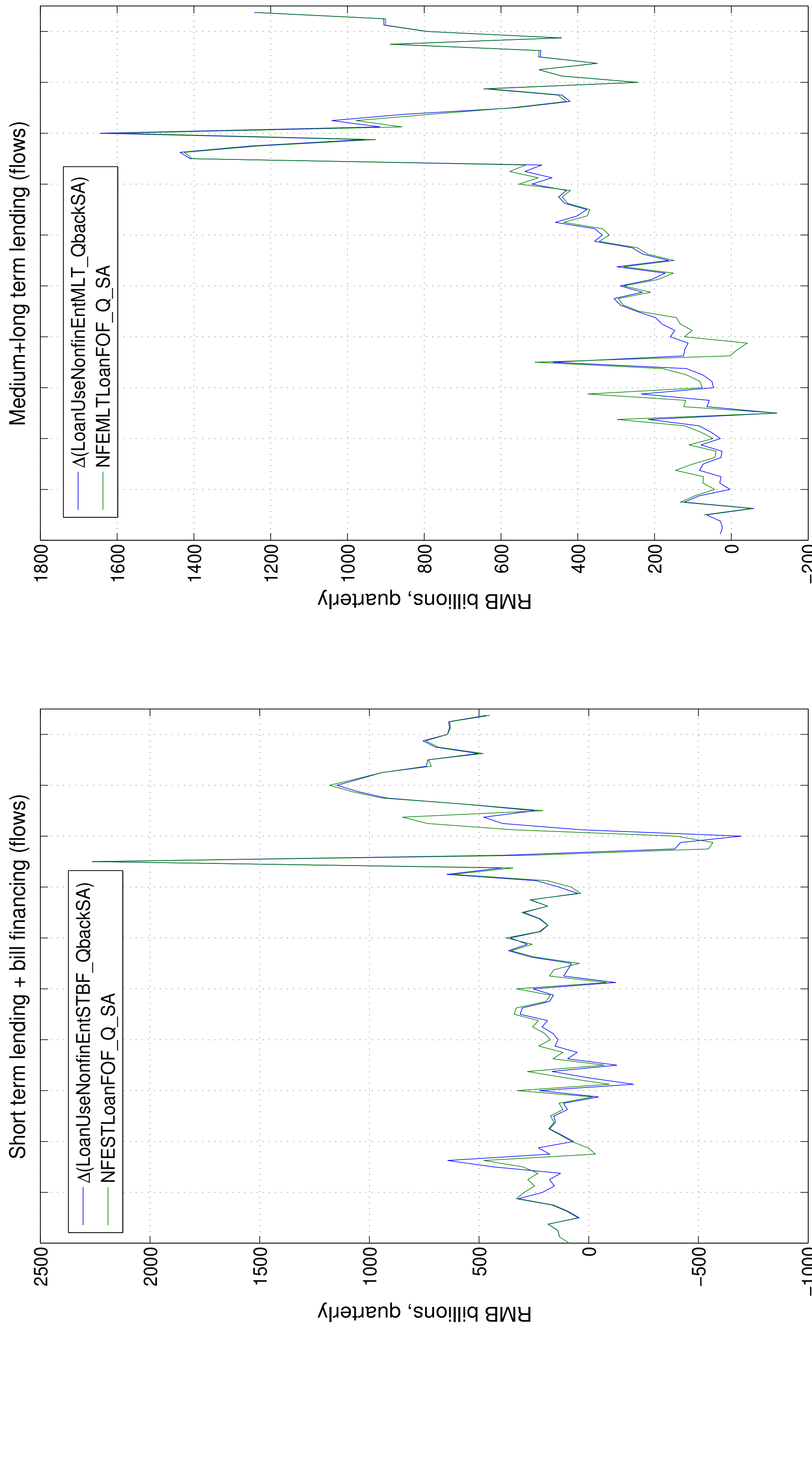

Figure 34. Alternative quarterly measures of flows of bank loans since 1994.

CEIC Ticker	Sector
CAAFWG	Non Financial Enterprise
CAAFYW	Financial Institution
CAAGBM	Government
CAAGEC	Household

Table 1. Subcomponents of annual gross fixed capital formation [CEIC ticker CAAFTQ]

CEIC Ticker	Sector	Starts	Assigned to
COMAA	State Owned	1980	SOE
COMAB	Collective Owned	1980	"Other Non-SOE"
COMAC	Individuals	1980	Private
COMAD	Joint Owned	1996	$\frac{2}{3}$rds SOE, $\frac{1}{3}$rd "Other Non-SOE"
COMAE	Share Holding	1996	"Other Non-SOE"
COMAF	Foreign Funded	1995	"Other Non-SOE"
COMAG	HK, Macau & Taiwan Funded	1995	"Other Non-SOE"
COMAH	Others	1996	Private

Table 2. Subcomponents of annual fixed assets investment [CEIC ticker COMA]

CEIC Ticker	**Sector**	**Assigned to**
COBDLV	State Owned Enterprise	SOE
COBDLW	Collective Enterprise	"Other Non-SOE"
COBDLX	Share Cooperative	"Other Non-SOE"
COBDLZ	Joint Enterprise: State Owned	SOE
COBDMA	Joint Enterprise: Collective	"Other Non-SOE"
COBDMB	Joint Enterprise: State Owned & Collective	$\frac{1}{2}$ SOE, $\frac{1}{2}$ "Other Non-SOE"
COBDMC	Joint Enterprise: Other	"Other Non-SOE"
COBDME	LLC: State Sole Proprietor	SOE
COBDMF	Limited Liability Company (LLC): Other	"Other Non-SOE"
COBDMG	Share Holding Limited Company	"Other Non-SOE"
COBDMI	Other Enterprise	Private
COBDMH	Private Enterprise	Private
COBDMJ	HK, Macau & Taiwan Funded Enterprise (HMT)	"Other Non-SOE"
COBDMO	Foreign Funded Enterprise	"Other Non-SOE"
COBDMT	Individual Business	Private

Table 3. Subcomponents of monthly (ytd) fixed assets investment [CEIC ticker COBDJU]

CEIC Ticker	**Sector**	**Sample period used**	**Frequency**
CGEAHA	Cash Income per Capita: ytd: Rural	1st quarters 98-00; 4th quarters 95-00; 01Q1-14Q3	Quarterly
CGEAA	Cash Income per Capita: ytd: Rural	1996Q1-2000Q1	Quarterly
CHAMBG	DPI per Capita: ytd: Urban: Avg	2002Q1-2014Q4	Quarterly
CGGA	Population	1949-2014	Annual
CGGD	Population: Urban	1949-2014	Annual
CGGE	Population: Rural	1949-2014	Annual

TABLE 4. Series used to interpolate annual disposable personal income and labor compensation

Description	CEIC ticker for **M0**	CEIC ticker for **M2**	**Start period**
$yoy_{t,m}$: Published monthly yoy %-change	CKSAAAA	CKSAACA	March 1997
$M_{t,m}$: Published monthly level	CKSAAA	CKSAAC	January 1997
$Q_{t,m}$: Published end-of-quarter level	CKAAA	CKAAC	March 1990

TABLE 5. Series used to construct M0 and M2

Description	CEIC ticker for **CPI**	CEIC ticker for **PPI**
$pp100_{t,m}$: Previous month = 100	CIAHJZ	CIAIEJ
$py100_{t,m}$: Previous year (12-months ago) = 100	CIEA	CIUA

TABLE 6. Series used to construct CPI and PPI

Description	**Coefficient**	**Standard Error**
Constant	1.52	1.35
CPI Inflation	-0.53	0.37
PPI Inflation	0.89*	0.27
GDP Deflator Inflation	0.36	0.56
*Statistically significant at 1% level.		

Table 7. Regression for investment price inflation (1990-2013) [R-squared=0.745]

Description	**Coefficient**	**Standard Error**
Constant	0.07	0.23
FAI Inflation	0.92*	0.028
*Statistically significant at 1% level.		

Table 8. Regression for GFCF price inflation (1990-2004) [R-squared=0.988]

Item	**Urban Ticker**	**Rural Ticker**
(1)Food+Beverage+Liq+Tobacco	CHABDD	CHABLQ
(2)Tobacco	CHAUAL	$CHABLQ\frac{CHAUAL}{CHABDD}$
(3)Liquor+Beverage	CHAUAM	$CHABLQ\frac{CHAUAM}{CHABDD}$
(4)Food	(1)-(2)-(3)	$CHABLQ(1-\frac{(CHAUAM+CHAUAL)}{CHABDD})$
(5)Clothing	CHABDJ	CHABLR
(6)Medicine+Medical Service	CHABDN	CHABLU
(7)Transport	CHAUBA	$CHABLV\frac{CHAUBA}{(CHAUBA+CHAUBB)}$
(8)Communication	CHAUBB	$CHABLV\frac{CHAUBB}{(CHAUBA+CHAUBB)}$
(9)Transport+Comm.	(7)+(8)	CHABLV
(10)REC*	CHABDP	CHABLW
(11)Recreational Articles	CHABDQ	$CHABLW\frac{CHABDQ}{CHABDP}$
(12)Education	CHAUBC	$CHABLW\frac{CHAUBC}{CHABDP}$
(13)Hh FAS**	CHABDL	CHABLT
(14)Hh Service	CHAUAZ	$CHABLT\frac{CHAUAZ}{CHABDL}$
(15)Hh FA	(14)-(15)	$CHABLT\frac{(CHABDL-CHAUAZ)}{CHABDL}$
(16)Miscellaneous	CHABDT	CHABLX
*Recreation, education and culture. **Facilities, articles and services.		

Table 9. CEIC tickers used to construct expenditure weights for CPI durables and CPI nondurables.

Item	**CEIC CPI Ticker**	**Exp. Wgt pre-2001**	**Exp. Wgt since 2001**
Food+Beverage excl Liquor	CIKI	Food	Food+$\frac{3}{5}$"Liqour + Beverage"
Liquor+Beverages	CIKIS	"Liquor+Beverage"	0
Liquor	CIAHGW	0	$\frac{2}{5}$"Liqour+Beverage"
Tobacco	CIKIR	Tobacco	Tobacco
Clothing	CIKK	Clothing	Clothing
Medical Instrument & Article	CIAHHA	$\frac{1}{8}$"Medicine+Medical Service"	0
Health Care	CIKMA	0	$\frac{1}{8}$"Medicine+Medical Service"
Transportation	CIKNA	$\frac{1}{4}$Transport	0
Fuel & Parts	CIAHHJ	0	$\frac{1}{4}$Transport
Newspaper & Magazine	CIKOE	$\frac{1}{5}$Recreational Articles	$\frac{1}{5}$Recreational Articles
Personal Article & Services	CIAHHD	0	Miscellaneous

TABLE 10. Annual CPI nondurable goods items and expenditure weights.

Item	CEIC CPI Ticker	Exp. Wgt pre-2001	Exp. Wgt since 2001
Hh Facility, Article & Mainten. Svc	CIKL	Hh FA	Hh FA
Medicines & Medical Appliances	CIKM	$\frac{3}{8}$"Medicine+Medical Service"	$\frac{3}{16}$"Medicine+Medical Service"
Health Care Appliance & Article	CIAHHB	0	$\frac{3}{16}$ "Medicine+Medical Service"
Transportation	CIKNA	$\frac{1}{3}$Transport	0
Transportation facilities	CIAHHI	0	$\frac{1}{3}$Transport
Communication	CIKNB	$\frac{1}{2}$"Communication"	0
Communication facilities	CIAHHN	0	$\frac{1}{2}$"Communication"
Educ.: Teaching Mat. & Refer. Book	CIKOB	$\frac{1}{20}$"Education"	$\frac{1}{20}$"Education"
Cult.&Recreat. Dur. Goods & Svc	CIKOA	$\frac{4}{5}$"Recreational Articles"	$\frac{4}{5}$"Recreational Articles"

TABLE 11. Annual CPI durable goods items and expenditure weights.

Item	CEIC PrevYr=100 Ticker	CEIC PrevMth = 100 Ticker
Food+Beverage excl Liquor	CIEB	CIAHKA
Tobacco+Liquor	CIEC#CIXC	CIAHKB
Clothing	CIED/CIAI	CIAHKC
Health Care	CIAHUJ#(CIAHKE*)#CIAK	CIAIGI
Fuel & Parts	CIAHUW#CIXM	CIAIGV
Newspaper & Magazine	CIAHVJ#CIXH/CIHM	CIAIHJ
Personal Article & Services	CIAHUP#CIXG	CIAIGO
Hh Facility, Article & Mainten. Svc	CIEE/CIAJ	CIAHKD
Health Care Appliance, Articles & Medical Inst.	CIAHUN#CIXF/CIAK	CIAIGM
Transportation facilities	CIAHUV#CIUAHE#CIAL	CIAIGU
Communication facilities	CIAHVB#CIXPC#CIEG	CIAIHA
Communication	CIAHVA#CIEG#(CIAHKF*)#CIAL	CIAIGZ
Educ.: Teaching Mat. & Refer. Book	CIAHVF#CIXH/CIHM	CIAIHE
Cult.&Recreat. Dur. Goods & Svc	CIXPB/CIXI	CIAHKG

*The ticker is a "previous month = 100" that is converted to a "previous year = 100" index.

TABLE 12. CEIC tickers used to construct interpolaters for annual item-level CPIs.

Item	CEIC CPI Ticker
Health Care Appliance, Articles & Medical Inst.	Average of CIAHHA and CIAHHB
Tobacco+Liquor (starting 2001)	CIAHGV
Tobacco+Liquor (pre-2001)	Tornqvist index of CIKIS [Liq.+Bev.] and CIKIR [Tobacco+Liq.+Art.]
*The expenditure weight for "liquor + beverages" is multiplied by 3/5 to proxy for the weight for liquour only.	

TABLE 13. Other item-level annual CPIs that are interpolated.

Item	**Expenditure weights**
Food+Beverage excl Liquor	"Food" + $\frac{3}{5}$"Liqour+Beverage"
Tobacco+Liquor	$\frac{2}{5}$"Liqour+Beverage"+Tobacco
Clothing	Clothing
Health Care	$\frac{1}{8}$"Medicine+Medical Service"
Personal Article & Services	$\frac{1}{2}$Miscellaneous
Transportation	$\frac{1}{4}$Transport
Newspaper & Magazine	$\frac{1}{5}$Recreational Articles

Table 14. Quarterly CPIs and expenditure weights used construct Tornqvist nondurables aggregate.

Item	**Expenditure weights**
Hh Facility, Article & Mainten. Svc	Hh FAS
Health Care Appliance, Articles & Medical Inst.	$\frac{3}{8}$Medicine+Medical Service
Transportation facilities	$\frac{1}{3}$Transportation
Communication facilities	Communication
Educ.: Teaching Mat. & Refer. Book	$\frac{1}{20}$Education
Cult.&Recreat. Dur. Goods & Svc	$\frac{4}{5}$Recreational Articles

Table 15. Quarterly CPIs and expenditure weights used construct Tornqvist durables aggregate.

References

Chang, C., K. Chen, D. F. Waggoner, and T. Zha (2015): "Trends and Cycles in China's Macroeconomy," *NBER Macroeconomics Annual*.

Fernald, J., I. Malkin, and M. Spiegel (2013): "On the Reliability of Chinese Output Figures," *Federal Reserve Bank of San Francisco Economic Letter*, 8.

Fernandez, R. B. (1981): "A Methodological Note on the Estimation of Time Series," *Review of Economics and Statistics*, 63(3), 471–476.

He, Q., P.-H. Leung, and T. T.-L. Chong (2013): "A Factor-Augmented VAR Analysis of the Monetary Policy in China," *China Economic Review*, 25, 88–104.

Holz, C. A. (2013a): "Appendix to The Quality of China's GDP Statistics," Unpublished Manuscript.

——— (2013b): "Chinese Statistics: Output Data," Stanford Center for International Development Working Paper No. 473.

——— (2013c): "The Quality of China's GDP Statistics," Stanford Center for International Development Working Paper No. 487.

——— (2014): "Monthly industrial output in China 1980–2012," *China Economic Review*, 28, 1–16.

NBS (2007): *Yearbooks: Historical Data on China's Gross Domestic Product 1952-2004*. National Bureau of Statistics, Beijing.

Ravn, M. O., and H. Uhlig (2002): "Notes On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations," *Review of Economics and Statistics*, 84(2), 371–380.

Waggoner, D. F., and T. Zha (1999): "Conditional Forecasts in Dynamic Multivariate Models," *Review of Economics and Statistics*, 81(4), 639–651.

Federal Reserve Bank of Atlanta; Federal Reserve Bank of Atlanta, Emory University, and NBER